REPORT RESUMES

ED 003 696 24

DEVELOPMENT OF SELF-INSTRUCTIONAL TECHNIQUES FOR TEACHING COMPLEX VISUAL DISCRIMINATION.

BY- GEIS, GEORGE L. ROTHENBERG, MARTIN
MICHIGAN UNIV., ANN ARBOR,CTR.RES.LEARN.AND TEACH.

REPORT NUMBER CRP-S-139 PUB DATE 65
EDRS PRICE MF-$0.18 HC-$4.20 105P.

DESCRIPTORS- *PROGRAMED INSTRUCTION, PROGRAMED MATERIALS, *TEACHING MACHINES, *VISUAL DISCRIMINATION, VISION TESTS, COLLEGE INSTRUCTION, STUDENT BEHAVIOR, *VISION, *AUTOINSTRUCTIONAL PROGRAMS, ANN ARBOR, MICHIGAN

THE SUBJECT OF THIS EXPERIMENTAL PROGRAM WAS THE RECOGNITION OF, OR CONCEPTUAL RESPONSE TO, COMPLEX VISUAL STIMULI BY COLLEGE STUDENTS. PROGRAMED INSTRUCTIONAL TECHNIQUES WERE USED IN THE SUBJECT CONTEXT TO DEVELOP PRELIMINARY SELF-TEACHING MATERIALS FOR LEARNING COMPLEX CONCEPTS. THESE PROGRAMING TECHNIQUES WERE IMPLEMENTED TO PRODUCE VISUAL STIMULI IN THE FORM OF SPECTROGRAPHIC SPEECH PATTERNS. CONCEPTUAL DISCRIMINATORY BEHAVIOR OF STUDENTS WAS RECORDED IN RESPONSE TO THE VISUAL STIMULI. SELF-INSTRUCTIONAL MATCHING TECHNIQUES WERE USED FOR BOTH REAL AND ABSTRACT SPEECH PATTERNS. THE PROGRAM PROVED EFFECTIVE IN ENHANCING OVERALL UNDERSTANDING OF COMPLEX STIMULI. IT APPEARED TO BE FLEXIBLE IN FORMAT AND DESIGN, AND ITS DISCRIMINATION POINTS (THE VISUAL CONTRASTS TO BE TAUGHT) COULD BE EMPIRICALLY DETERMINED BY THE PARTICIPATING STUDENTS. (JH)

ED003696

DEVELOPMENT OF SELF-INSTRUCTIONAL TECHNIQUES FOR TEACHING COMPLEX VISUAL DISCRIMINATION

Cooperative Research Project No. S139

George L. Geis

Martin Rothenberg

Center for Research on Learning and Teaching

The University of Michigan

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

DEVELOPMENT OF SELF-INSTRUCTIONAL TECHNIQUES FOR TEACHING COMPLEX VISUAL DISCRIMINATION

Cooperative Research Project No. S139

George L. Geis

Martin Rothenberg

Center for Research on Learning and Teaching

The University of Michigan

Ann Arbor, Michigan

1965

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education and Welfare.

ED 003696

STAFF

Project Director: Stanford C. Ericksen

Co-investigators: George L. Geis
Martin Rothenberg

Research Assistants: Kirk Wheeler
William Jacobs

Table of Contents

Introduction.. 1

Generation of Stimuli... 7

Experimental Design.. 15

Pretest.. 16

Programmed Materials... 26

Revision of Program.. 33

Control Groups.. 38

A Comparison of Training Methodologies........................... 39

Transfer... 43

Conclusions and Implications....................................... 45

Figures

Tables

References

Appendix A

Appendix B

1. Introduction

Much of learned behavior, most of human behavior, and almost every bit of human behavior which has been acquired through formal instruction, may be termed discriminative responding. The emission of a discriminative response is controlled, to a large extent, by a specific set of environmental conditions, or, stimuli; if these stimuli are not present the response is unlikely to occur. The verbal response RED (color) is controlled, for the most part, by a wave length of the object being observed; the color response BLUE, is also controlled by wave length, but, of course, by a different wave length. Were the environment simple, a specific, invariant stimulus would control a specific discriminative response, and teaching discriminations would be straightforward. But the stimulus situation is not invariant; two blues are similar but not identical. A response must be generalized to a whole population of stimuli. The aim of instruction is often the development of distinctive responses to sets or classes of objects rather than to a pair of single items--the development of concepts rather than simple discriminations.

Furthermore, most concepts are not based upon a single dimension of stimuli but upon multiple stimulus dimensions; a class is defined by, and distinguished from other classes on the basis of, not one but several cues. Most conceptual responses, then, must be controlled by patterns of cues which define the class or concept. This is the kind of behavior and the kind of control that academic instruction is in great part attempting to produce though the specific behaviors may range from map reading to EEG reading, from recognizing a painting as one from Picasso's blue period to recognizing an infant as an Rh negative baby.

Since the development of conceptual behavior controlled by multiple dimensions of complex stimuli is a major function of education, one would expect to find numerous and systematic series of experimental studies probing various aspects of the area. A search of the literature of Psychology and Education fails to reveal anything like the treasury of accumulated experimental findings and knowledge that the designer of instructional materials might hope and expect to find. "Concept Formation" and "Abstraction" are headings for numerous listings each year in the Psychological Abstracts; "Concept Learning" appears repeatedly as the title of a chapter in Psychology books and sometimes as the title of the whole book (e.g., Hunt, 1962). Experimental studies of recognition of, or conceptual responding to, complex stimuli are, however, quite rare. One exception, of course, is the work of Heidbreder beginning in 1946. But even these studies are not directly applicable to real world instruction--to the problems of teaching complex concepts. A few truly relevant studies can be found in the literature of human engineering and military psychology (e.g., Dawkins, 1964; Weisz and McElroy, 1964). As often happens, these applied studies lack the generality one might hope for.

With a dearth of guidelines the instructional designer may turn to very basic principles in psychology and extrapolate from these. The psychology of learning would suggest that teaching complex concepts, at least efficiently teaching them, would involve presenting numerous examples of these concepts. Furthermore, probably, the learner should be required to respond to each example, classifying it as a positive or negative example of the concept. (The sizable literature on matching to sample procedures has recently been examined by one of the investigators [see Geis and Spenser, 1965] and seems immediately relevant to the present task.)

A study of the application of these few principles to a real-life learning situation might well use, as the teaching medium, programmed instruction. It has been repeatedly pointed out by programmers that programmed instruction can be both an instrument for teaching and a tool for research. Unfortunately, this Janus-like role is rarely demonstrated. One aim of the present study was to develop teaching materials. The careful specification of instructional goals, the iterative student testing, etc. which characterize programmed instruction recommended it as the teaching medium. A further concern of the present study was an examination of the role of examples in teaching. Programmed instruction provides the opportunity for continual monitoring of student performance, a format which allows for precise manipulation of relevant variables, etc. And so it appeared to be a desirable research tool as well.

With a method arrived at which provides the chance to investigate a variety of problems in the applied science of instruction, content must be considered. Reading visible speech patterns produced by spectrographic analysis is a behavior that illustrates the conceptual responding described earlier. The stimuli are complex, with the visual display providing information on three parameters of the acoustic speech wave (i.e., frequency, duration, intensity). Correct responding requires distinguishing one speech pattern from another when the patterns represent different phonemes but classifying two speech patterns in the same way when a single phoneme is being represented in both. Few people can read these patterns. This means that an ample supply of naive subjects can be obtained. And furthermore, any effective teaching materials developed from such studies might well prove to be a boon to students in speech, linguistics and communication sciences, who in general share the layman's naivete in regard to visible speech reading.

A not incidental consideration in choosing visible speech patterns as the stimuli to be used was the fact that one of the investigators had earlier developed a preliminary program to teach visible speech reading (Rothenberg, 1964). Thus, we had some familiarity with the procedures of selecting and producing the stimuli, with the preliminary analysis of some of the difficulties in learning these discriminations, and with the problems and techniques of programming these materials.

A precise case history of the experimental development of programmed materials can be exciting, and provide a rich store of observations both about programming and learning (see, for example, Hively, 1962). The present study was designed in part to produce such a document of program evolution. It was designed, in addition, to have a more traditionally experimental flavor with variables being varied singly and the results being observed and recorded.

The study was designed to provide materials for several case histories. For example, a detailed record of the process of developing and selecting the stimulus materials was kept and is reported. The specification of, and the generation of, the teaching and testing materials was a major aim of the research. The criteria for acceptance of such materials were numerous and demanding; the mechanical procedures needed to generate acceptable stimuli were elaborate and time consuming. The considerable effort expended in the selection of and generation of stimuli (i.e., of examples to be used as teaching materials) may be in part a function of the content itself. But the effort also reflects the investigators' extreme concern with precise control over the specific materials to be used in teaching. Our experience in preparing other programmed materials indicates that the amount of effort expended on the generation of materials has not been unwarranted but was unusual or atypical when compared with the developmental procedure for most educational materials. Indeed the effort is a demonstration of and a prescription for, the careful generation of stimuli for teaching.

A part of the study is designed to illustrate the relative effectiveness of two approaches to the definition of instructional goals. The specification in fine grain detail of these teaching goals is, in fact, the specification of each discrimination and concept to be taught. The present study generated such statements of objectives both through a careful, *a priori* logical analysis and empirically through the use of student data. The latter data were obtained from tests designed to pinpoint confusions, or incorrect generalizations, that the students were making before training. A rerun of these tests during and after training provided evidence of the degree to which the intervening teaching had been effective. The former method of determining specific teaching goals (i.e., through logical analysis) is almost exclusively the way teaching materials are designed today. Only upon rare occasions is the design based upon empirical data. Both procedures provide not only the definition of the set of concepts to be taught but also an indication of the relative difficulty of each concept (thereby indicating the amount of instructional effort needed to teach each concept).

One aim of the study, then, was the generation of stimuli; another aim was the exploration of alternate ways of generating the goals of instruction. Still another aim involved the examination of the means to those ends. Once the instructional mode has been decided upon--in this case, programmed instruction--the language of instruction must be determined and the particular materials selected. In the case of a program, especially one which might be termed non-verbal, the number, types, and placement of examples is of immediate concern. The role of examples is just beginning to develop as a problem area in instruction. A preliminary exposition of our own interest in this area appears elsewhere in the report (see Appendix B) and will not be detailed here. Suffice it to say that one of the first major decisions to be made in the selection of examples is whether to use real life examples or abstractions, models, etc. Often in the development of educational materials the decision is simply not made, if by decision is meant a rational judgment based upon some weighing of evidence. A drawing is placed in a textbook because the instructor thinks that it is clearer than a photograph; a photograph is used because the author thinks it will add interest; a model is used in class because it is available. But the choice of examples of the concepts being taught need not always be adventitious. The present study attempts to explore the relative advantages of abstract and real patterns in teaching the recognition of both real and abstract patterns (i.e., the effect of pattern on original learning and transfer).

A rather fine point of programming methodology was also investigated in the present study. The teaching programs involved a matching-to-sample format: in each teaching frame the student was required to choose from a pair of alternatives the one which matched a sample stimulus. The effect of varying number of alternatives (i.e., number of examples in a frame) was examined by providing one group with a program which had more examples per frame than did the other programs.

No discussion of teaching visible speech would be complete without the mention of at least two works--one from the past and one which seems to point to the future. The classic work Visible Speech (Potter, 1947) is a textbook designed to teach the discriminations being taught in this study (and more). The emphasis in the book was on teaching the deaf real time visible speech reading so that in many respects our efforts and Potter's are not comparable, quite aside from the much greater magnitude of his efforts. But both are attempts to teach and in less than twenty years the textbook has come to seem an especially archaic way of teaching these discriminations. Programmed instruction appears a much more appropriate mode of instruction. Casting a shadow over efforts to improve the teaching of visible speech reading, such as the present study, is the continuing work to produce computer-reading of speech, represented, for example, by the work of Uhr and Vossler (1961). While computer reading is still more dream than reality, any system of teaching visible speech to human beings may become outmoded. What will remain useful, however, is any light that can be cast through studies like this one on the complex process of human discrimination learning and on the problems of optimizing formal instruction designed to produce such learning.

2. Generation of Stimuli

In order to generate the stimuli used in the study it was necessary to: (1) construct a consistent theoretical model which would provide the basis for intelligent selection of the speech sounds to be used, (2) determine the optimal means for converting the acoustical signals into visible speech records, and (3) design and produce the idealized speech patterns. What follows is a description of the rationale for sound selection as well as a description of the sound samples, the production of spectrograms, and the development of idealized patterns.

Rationale for Sound Selection.

The task of this experiment was to study programed methodology for teaching discrimination between speech spectrogram patterns made from speech sounds belonging to different sound types, or phonemes, of English. For experimental purposes, it was necessary to delimit this task somewhat further. The major criteria used in choosing a set of speech sounds from which spectrograms were made were these:

1. The sounds must be clearly related to one, and only one of those speech sound classes in English used in this study.
2. Sounds from different speech sound categories must be separable on the basis of their spectrograms by an observer not using measuring devices or statistical techniques.
3. There should be a significant number of sample pairs similar enough to produce some discrimination problems in untrained observers.
4. The set of speech sound classes represented, although not necessarily complete, should be large enough and contain enough variation within each class so that the material taught will make a significant contribution to the learner's ability to discriminate between speech spectrogram patterns.

5. In connection with the construction of abstract or idealized patterns a final criterion must be added. The set of sound samples selected must be such that their spectrograms can be at least roughly separated on the basis of a well-defined set of abstract concepts existing today in the field of instrumental speech analysis. (A description of these concepts appears later in this report.)

Description of Speech Samples Selected.

Applying these criteria, a set of sound samples was selected. The speech sound classes being employed were restricted to a subset of the continuant English sound classes. A list of the phoneme-related continuant speech sound classes is given in Table I. with the subset employed in this study indicated. The specific reasons for not using some entries in the Table will not be detailed here; however the criteria described above were used in the selection.

The sound samples themselves were short, approximately 60 millisecond segments extracted from a representative portion of the occurrence of the particular speech sound type in a consonant-vowel-consonant word. These words were spoken on tape by one male, native midwestern English speaking, phonetically trained person. The words were spoken in the context "Say the word ______ again."[1] (Lehiste and Peterson, 1959) Each sound sample was listened to in context, and its spectrogram examined, to make sure that it was a clear, unambiguous representation of the appropriate sound type. The 60 millisecond segment most closely approximating a representative articulatory steady state was selected by a trained observer (M.R.). Where there was latitude for choice, the most centralized segment was selected.

The words containing the samples were selected so as to give a wide dispersion of phonetic environments for the examples in each class. A list of 121 CVC words that would produce a minimum of 14 examples of each speech sound class was compiled.

---

[1] We wish to thank Dr. Gordon Peterson, Director of the Communication Sciences Laboratory (University of Michigan) for making these tapes available.

Development of Materials: Speech Spectrograms

The properties of a speech spectrogram can vary widely depending on the spectrograph machine used and the particular adjustments of that machine. To an experienced observer, many of the variations possible are secondary, since he automatically compensates, or at least recognizes, those qualities of the spectrograms due to the type of machine and its adjustment. In particular, many machines are constructed and adjusted so as to make less visible, or not visible, many of the features of the acoustic spectrum which are believed to be less significant in the identification of speech sounds. Conversely, machines are designed to emphasize those features which are thought to be most significant by the designer.

For this study it was decided to use a speech spectrograph machine providing spectrograms with minimal distortion. "Distortion," here, means deviance from that display which would be associated with the theoretical notion of the speech spectrogram. Spectrograms were examined from a number of machines, including three at the University. The machine finally chosen was a modified Kay Sonograph at the General Dynamics Laboratories, Rochester, New York, which retained the primary features of the common Kay machine, including the fairly standard Kay wide-band (300 cps) filter.[2] However, this machine contains a significant modification in that the marking voltage, and therefore, roughly the density or darkness of the display, is proportional to the logarithm of the spectral intensity. This modification permits an expansion of the range of intensities visible on the spectrogram. The modified spectrograms can present an intensity range of over 30 db, compared to about 12 db using the original linear display. In addition, after considerable experimentation, it was decided that the spectrograms should be produced at a gain setting that saturated on the very darkest parts of the display (i.e., a gain setting about 6 db higher than the maximum setting for no saturation). The total range available is, then, about 40 db, which is sufficient to display all except the lowest intensity speech sounds at the same gain settings as the high intensity speech sounds. In this way variations of overall intensity between speech sound samples are preserved without sacrificing much information at the highest or lowest intensity levels.

---

[2] The cooperation of Robert Houde and staff at this laboratory is gratefully acknowledged.

Speech spectrograms are generally produced with some increased amplification at the higher frequencies to compensate for the decreased energies generally present in the upper part of the speech spectrogram. It was decided to use a high frequency emphasis of 6 db per octave, which makes visible most of the spectral features of interest in the high frequency range used, but, in most speech sounds, still retains a relative lightness in the display for the weaker high frequency energy concentrations.

The modified Kay Sonograph at General Dynamics, though highly suitable for this work, produced a number of idiosyncratic deviations in its display, and it was necessary to go through a considerable number of trial spectrograms using various settings for gain, filtering, needle pressure, etc., before spectrograms were obtained which were suitable. The problem of obtaining an acceptable set of spectrograms caused considerable difficulty in the project. Many of the minor deviations in the display which are acceptable to the trained user might be misleading to the beginner. It was decided that, for the purposes of this study, it would be desirable to eliminate, as far as possible, all pattern deviations traceable to such factors as noise or non-optimal adjustment of the equipment. In an actual training situation some of these factors might be introduced at a later stage in training.

The time scale of the spectrograms used was 5 inches per second. The frequency scale is almost linear and covers about 1500 cps in one inch. The actual spectrogram samples used as stimuli in the study were vertical strips, 5/16ths of an inch by 3½ inches. A sample spectrogram is shown in Figure 1. with the frequency and time calibration noted.

Development of a Set of Idealized Patterns

To provide the rules for constructing a set of idealized patterns from measurements of the real spectrogram samples it was necessary to develop a model, or a set of concepts, for the information-bearing acoustic parameters of speech. The model had to be sufficiently strong to provide enough information to identify the sound class of a great majority of the samples used in the study. Also, it had to reflect the type of concepts or idealizations that might be presented to a beginning student in the field of instrumental speech analysis. The idealized patterns constructed from this model will retain those features and only those features that are represented in the model. In more detail our task was to:

1. Identify a set of parameters that are believed to be significantly correlated with the differences between 2 or more sound classes. Some parameters may be included even if their correlation or significance is not known with certainty.
2. Specify in detail how these parameters are to be measured in real spectrograms.
3. Specify in detail how the resulting parameter values are to be used in the construction of a set of idealized or abstracted patterns.

The idealized model for vowels (/u/, /ʊ/, /ɔ/, /a/, /I/, /ɛ/, /i/), semi-vowels (/l/, /r/) and nasals (/m/, /n/) were specified by the three lowest resonances or formants, including the frequency and rate of change of each formant, and by their voice frequency. All three formants were specified. A glottal fricative (/h/) was specified by whichever of its three lowest resonances or formants were visible on the type of speech spectrograph display being considered. Unvoiced fricatives and sibilants (/θ/, /f/, /ʃ/, /s/) were specified by the lower boundary of the lowest primary energy concentration, its frequency and rate of change, and a 2-valued strength parameter, strong or weak. The voiced sibilant (/z/) was represented, likewise, by the lower boundary of the lowest strong energy concentration, its frequency and rate of change, and, in addition, by the pitch or voicing rate. (Since /z/ has no "weak" counterpart in the sound types considered in this study it was not necessary to specify a strength parameter.)

The required measurements of the spectrogram samples were made by one of the co-investigators (MR) using a procedure that was as algorithmic as possible. There were relatively few cases in which a judgment was used which could not conceivably have been mechanized. Many of the problems involved in the determination of such parameters as formant frequency and voicing frequency from speech spectrograms have been widely discussed in the literature. Details of the measurement procedure will not be discussed here, except to note that the average voicing rate in the 60 millisecond period of the sound sample was determined rather crudely by matching 4 vertical grids, with different spacings, to the vertical striations on the spectrogram that are related to the voicing pulses. Since voicing rate is thought to have little correlation with any of the boundaries between the speech sound classes employed, this crude classification appears to be adequate.

Figure 1 contains three sample idealized patterns along with the spectrogram samples from which they were derived. The striated and randomly patterned areas were cut from commercially available transparent adhesive-backed materials. They were selected and laid out to represent the measured parameters of the model in a manner suggestive of the role of these parameters in the actual spectrogram.

We have made use of the following patterns:

1. Artype patterns No. HR4044, HR4045, HR4046 and HR4047, which are striated at 16, 20, 24 and 28 lines/inch respectively. These spacings are equivalent to voicing frequencies of 80, 100, 120 and 140 cps respectively and cover a range adequate for the sounds being idealized.

2. Artype patterns No. HR4081 and HR4083, which have randomly marked patterns suggestive of the pattern produced by filter gaussian noise on a spectrogram. The two patterns differ primarily in density and will be referred to as the light and dark noise patterns respectively.

The idealized patterns were mounted on a silk-screened gray paper which resembles the spectrogram paper used.

We have used the following rules in constructing the idealized patterns:

1. The narrow formant bars indicating the resonances of the vowels, nasals, and glottal fricatives were represented by a bar slightly more than 1/4 inch wide, made of striated material in the case of the voiced sounds, (with the number of striations/inch similar to the number of striations in the spectrogram), and made of light noise material in the case of a glottal fricative.

2. The non-voiced fricatives and sibilants were represented by a single expanse of noise pattern extending from the lower boundary measured up to the top of the pattern. For /s/ and /ʃ/ a dark noise pattern was used; for /f/ and /θ/ a light noise pattern was used.

3. The voiced sibilant /z/ was represented by dark noise (as with its unvoiced counterpart /s/) overlaid with a striated pattern of the appropriate spacing. A narrow (1/8 inch) strip of similar striated material on the bottom margin simulated the "voice bar."

4. "Voicing" alone is simulated by a narrow (1/8 inch) strip of striated material at the lower margin.

5. "Silence" is represented by no pattern.

6. All idealized patterns have a horizontal line drawn at 190 cps to represent the fact that the real spectrograms do not show the lowest 190 cps of the spectrum. No pattern material is shown below this line.

Both the real and abstract patterns were photographically reproduced in actual size to provide multiple, durable copies of the stimuli. A matte photographic paper was used to simulate the texture of the original spectrogram. Reproductions were compared closely with the originals, and it was concluded that little detail, if any, had been lost by photographing, although some extremely light details were lightened further. Copies of each pattern were coded, catalogued, and kept in a pattern file.

3. Experimental Design

Because this study was designed to probe and to explore, no serious attempt was made to conform with traditional experimental design. Consequently, a natural sequence for this report seemed to be suggested by the design itself rather than by the more conventional topic arrangement (i.e., Method, Results, Discussion.) Therefore, the report is organized so that each stage of the experiment appears as a separate section and each section includes the relevant procedures, results, and discussion.

An overview of the entire design can be seen in Figure 2. All Ss went through the PRETEST confusion matrix, using either REAL or ABSTRACT stimuli. Then S followed either the PROGRAM or CONTROL route. PROGRAM Ss worked through a 120 frame teaching program in which the stimuli matched those of the pretest. CONTROL Ss continued to run through the pretest confusion matrix in three successive sessions. Control Ss with FEEDBACK could discover the answer to each item after answering that item while subjects in the NO FEEDBACK groups merely went through the "pretest" three more times.

All Ss were then returned to the pretest procedure; this was now the POSTTEST, designed to measure the effect of the intervening activities. Finally, Ss were run through the TRANSFER test which was simply the other group's (ABSTRACT or REAL) pretest.

All Ss were college students and were paid at the usual rate for experimental subjects. Ss were always run in the laboratories at the Center, and, whenever possible, were run on successive days.

4. The Pretest

One of the primary goals in this study was to accumulate data concerning the discrimination problems inherent in the stimuli and stimuli classes used in this study. A confusion matrix was constructed from generalization data obtained from college students who were unfamiliar with speech spectrogram patterns. The matrix was used both as a design instrument for the construction of some program materials and as a measure of learning for students going through the test programs. The confusion matrix is an attempt to objectify and quantify that which good teachers have often used when designing teaching materials, namely, observations of confusions and difficulties the naive student demonstrates when he encounters a new subject. Since little is known about deriving teaching materials from confusion matrix data, an important part of this study was to make an initial exploration into methods for extracting rules from the matrix that could be used for designing teaching materials.

Procedure. The following discussion applies both to the REAL and the ABSTRACT patterns.

From the pattern file approximately[1] fourteen samples from each of the nineteen stimuli classes were selected randomly to be used as a set of test stimuli. Each sample was mounted behind a die-cut slot in a matte-black cardboard card. The cards were then attached to the pages in 3" x 5" photograph display books. Each book contained twenty-five pages.

---

[1] A few labeling errors led to a slight imbalance between classes. Since the imbalances were slight, the number of samples from each class was allowed to vary one or two from the average of fourteen.

Canonical Forms. A significant problem was the selection of a set of "representative" instances, one from each stimulus class (i.e., each phoneme class) to act as "canonical forms". (In the experiment the subject was to match samples in the books against this set of canonical forms, and to select the one canonical form that the sample matched closest.) The following selection criteria for canonical forms were used:

1. All vowels were taken from those words in which they occurred after an /h/. In this position the constraints of articulatory dynamics are least, since the vowel can be formed during the articulation of the preceding consonant.
2. All consonants are taken from words in which they occurred in the initial position before an /a/ sound. In this position the tongue tip and tongue blade, and the lips, are minimally restricted by the movements that have to be made for the production of the following vowel.

When more than one sample occurred in the proper phonetic environment, the sample that seemed clearest was chosen. The ABSTRACT canonical forms were the abstract versions of the REAL canonical forms.

The canonical forms were mounted on 1"x 5$\frac{3}{8}$" sheet metal strips. The strips had been previously covered with black cloth tape. The patterns were mounted under a slot in a black matte cardboard mask as in the case of the samples, and then mounted on the metal strips. The top of the strip was bent up at an angle to provide a convenient tab by which the strips could be handled. A photograph of the resulting canonical forms is shown in Figure 3.

The experimental set-up for obtaining the confusion matrix data is illustrated in Figure 4. The fourteen strips containing the canonical form patterns were arranged on a horizontal magnetic bar mounted on the top portion of the sloping surface in front of the student. The samples, one book at a time, were placed immediately beneath the canonical forms. To the right was clipped a sheet on which the subject noted his answers.

A copy of the instructions to the subject can be found in Appendix A. S looked at one sample at a time, in the order presented in the book, found the closest matching canonical form and wrote down its number on the answer sheet. S was free to remove the canonical form strips from the magnetic rack above and compare them to the sample. No restrictions were made on the subject's strategy, though the experimental setup was arranged to make it convenient for the subject to remove two strips at one time (one with either hand) and lay them on each side of the unknown.

The canonical forms were initially arranged so that those patterns that are most obviously the same were placed in rough proximity and numbered in that order. It was found that many subjects, at the beginning of the test, rearranged the order of the canonical form strips on the magnetized rack.

Results. The raw data consisted of S's matching responses. For each experimental group, these data were entered on IBM cards and computer printouts were obtained in the form of confusion matrices. The axes of these matrices consisted of patterns sent (i.e., samples) and patterns received (i.e., canonical form S matched to sample). Tables I and II show typical pretest matrices.[1]

The ordering of the stimuli classes along the axes of the matrices was determined by trial and error from the original confusion matrix data for the REAL patterns. The order of the stimuli classes was arranged to obtain a matrix in which the larger numbers clustered along the diagonal. The possible relationships of such clusters to the teaching problem will be discussed below.

---

[1] The tables shown are from preliminary runs with larger Ns than were used in the experimental treatments. For these tables the N = 16.

The fact that a body of confusion data can be organized so that there is a large degree of clustering along the diagonal indicates an inherent organization among the patterns of confusions. In this respect it is interesting to examine the difference between the matrices for the REAL and ABSTRACT patterns. As can be seen in the tables, the REAL matrix is organized more tightly around the diagonal than the ABSTRACT matrix. This difference is not merely due to the arrangement of the stimuli along the axis; little improvement could be obtained by rearranging the pattern sequence in the ABSTRACT matrices. It will be seen from the discussion below that it was easier to use the REAL confusion matrix data to generate a programmed teaching sequence designed to remove the underlying discrimination problems.

As might be expected, the total percentage of errors was greater for the REAL case than for the ABSTRACT case. As can be seen in Table IV , when REAL subgroups were combined, the pretest error was 20% greater than for Ss in the ABSTRACT groups. The error data verified an original hypothesis that the patterns and pattern classes contained a great number of easily recognizable differentiating characteristics, but also presented a sizeable number of discrimination problems for the untrained observer. The pretest confusion matrices indicated that the discriminability of the patterns was somewhat greater than predicted, with many classes showing very few confusions. This conclusion is of considerable significance in the organization of teaching materials: there are apparently numerous dimensions that need not be trained to any significant extent, they need only be "verified" for the student; he merely needs to know that his original discrimination is correct. It would seem that teaching materials organized without reference to a detailed analysis of the inherent discrimination problems would tend to be excessively inefficient in teaching those dimensions that are easily discriminable.

Given a confusion matrix with a high degree of organization, such as the matrix for the REAL patterns, what can be said about the underlying patterns of confusion? In this study an effort was made to analyze the matrix for the REAL patterns (Table II) with the hope of decomposing the total matrix into a set of 'submatrices', each submatrix indicating some significant pattern of confusions, and such that the totality of the submatrices could be said to cover most of the discrimination problems indicated by the matrix. The term 'submatrix' is not used here in its strict mathematical sense; this term is meant only to denote a pattern of related confusions indicated on the matrix. In a highly organized confusion matrix, most, though not all, significant patterns of confusion are indicated by high entries within a rectangular area on the confusion matrix (a submatrix). (The terms 'pattern' or 'subpattern' might be more descriptive than 'submatrix', however it was felt that these terms would conflict with the frequent use of the word 'pattern' to describe the stimuli.)

An attempt was made to define a set of submatrix types, such that each distinct type represented a distinct type of confusion problem, and therefore indicated the necessity for a distinct teaching method. An analysis of the confusion matrix for the REAL patterns disclosed that there were a relatively small number of submatrix types that could cover most of the confusions indicated on the matrix. These matrix types will be discussed below and illustrated from the simplified confusion matrix for the REAL patterns shown in Table V. In this matrix all entries of less than 16 (the number of subjects) are eliminated. Thus the matrix in Table V shows most of the important patterns of confusion but not of the scattered smaller numbers which may have resulted from only one or two deviant patterns. The diagonal elements (the correct decisions) are also not shown. In each case some speculations will be presented concerning the relationship of the submatrix type to the actual causes of confusion in the patterns used in this study.

1. There are a number of examples in Table V of mutual confusions between two pattern classes. When the mutually confused classes can be brought adjacent to one another, the result is a "square" on the diagonal of the confusion matrix. Some pairs that are mutually confused are /u/ and /l/, and /f/ and /θ/. Wherever there was a mutual confusion in which the errors in both directions were approximately balanced, an analysis of the patterns involved indicated that the patterns in both classes were acoustically similar, and the acoustic dimensions tended to overlap. Often, it was theoretically possible to discriminate between the pattern classes but some complex higher order rule was required.

2. In a great many cases there were confusions that existed only in one direction, i.e., Class A was sometimes called Class B, but rarely the reverse. When the confused classes are adjacent in the confusion matrix the result is a "triangle" on the diagonal. There seem to be a number of different causes for this type of error. In some cases it seemed as though an incorrect dimension, or, more precisely, a partially incorrect dimension was being used. This seemed to have occurred in the discrimination between /n/ and /h/; the dimension of dark versus light would distinguish between these two in most cases, but occassional examples of /n/ could be found which were as light as the average /h/. (A more reliable indicator than darkness would be vertical striation.) A slightly different problem seemed to arise in the discrimination between /z/ and /s/. In this case there were a number of /z/ samples in which the difference in quality between the /z/ and /s/ was marginal, though the indicators of voicing, some slight striation and a small trace of a "voice bar" at the bottom margin of the pattern, could be found in /z/. It is likely that a number of the /z/ patterns were called /s/ because the attention of the subject was not focused on these slight distinguishing characteristics. The critical point here was that even a

slight trace of a "voice bar" is enough to label a pattern a /s/. The significant characteristic of this distinguishing feature seems to be that it does not require a discrimination to be made along some continum. A similar but opposite situation occured in the one directional confusion of /#/ judged as /+/. Though "blank" patterns (i.e., /#/ pattern type) would seemingly be easy to discriminate apparently some subjects hypothesized that a trace of pattern —a smudge or speck no matter how small —would be sufficient to disqualify a pattern from being a /#/ and would label it a /+/ (a voice bar). Here again there was no continum involved in the discrimination required.

A somewhat puzzling confusion occured when /r/ was occasionally judged as the rather unsimilar /+/. An analysis of the errors showed that a number of the /r/ samples were occasionally called /+/ and that the errors are not due to just one subject. Also, no distinguishing characteristics could be found in the confused /r/ samples. (A possible explanation for the confusion between /r/ and /+/ might be found in the explanation for the one way confusions involving /l/ below.)

3. The remaining submatrix types to be discussed will stem from one way confusions involving more than two pattern types. Perhaps the more obvious of the submatrices involving more than two pattern types is the 3 x 3 matrix formed by the types /ɔ/, /a/, and /U/. As in the case of the 2 x 2 matrices, the pattern types involved in this 3 x 3 matrix arise from acoustically similar sounds, and the patterns must be differentiated by a sometimes complex weighting of judgments concerning more than one parameter.

4. No attempt will be made here to discuss exhaustively the different types of unbalanced submatrices involving more than two pattern types. However, there are a number of relatively simple unbalanced matrices which do occur and which appear interesting. For example, a case in which a single pattern type is confused with two or more other pattern types but

in which no (or few) confusions in the reverse direction occur. This type of confusion can, in turn, be broken up into two cases. The first case is that in which the various incorrect responses are themselves mutually confused (and, therefore, in a well-organized confusion matrix they will occur adjacent to one another). One such example is the case of /i/ being incorrectly labled either /n/ or /m/. This type of confusion tends to occur as a horizontal bar on a well-organized confusion matrix. There is some justification for considering this type of confusion to be essentially similar to the one way confusion between just two pattern types. The essential point here is that the incorrectly chosen pattern types are themselves very similar and often confused, and, in this case, can be considered as one class.

The second type of confusion in which one pattern type gets placed incorrectly into two or more different categories occurs in the case of the /l/. The /l/ is judged as being either /u/, /r/, /n/, /m/, or /s/. The /l/ vs. /u/ confusion can be considered a mutual confusion since it occurs in both directions, and indeed, /l/ and /u/ are acoustically similar. The same could be said for /l/ versus /r/. However, the confusions with [/n/, /m/] (they can be placed together in one class, as discussed above) and the confusions with /s/, are one way confusions. Likewise, the /l/ class is not extremely similar acoustically to the pair [/n/, /m/] or to /s/. The only explanation we can give for these confusions is that because of the variety of the different manifestations of an /l/, and the general lightness of most of the manifestations, it was difficult for many of the subjects to have an adequate a priori concept of set inclusion for the /l/ pattern type. Assuming this to be true, it is easy to understand how individual instances of an /l/, when not readily identified with the correct pattern type, and when not extremely similar to any other one pattern type, would be placed semi-randomly into other classes of partial similarity. In other words, it might be hypothesized that if one pattern is incorrectly judged as a number of other patterns, and these incorrect categories are themselves somewhat different and not mutually confused, the problem is not one of a deficient discrimination between pattern types but a problem of an inadequate concept of set inclusion for the pattern type being mislabled. On a well-organized confusion matrix this type of confusion would occur as a horizontal bar that is somewhat spread out in its horizontal dimension.

5. Confusions have also occurred which show up on the confusion matrix as vertical bars, as opposed to the horizontal bars discussed above. The one clear example is the case in which /h/, /f/, and /θ/, are at times mislabled as /#/. It is important to note that /h/, /f/, and /θ/ are confused among themselves to some extent, i.e., that /h/, /f/, and /θ/ apparently share some common dimensions. Thus, a hypothesis could be raised that it is variations along some one common dimension that caused examples of /h/, /f/, and /θ/ to be mislabled /#/. This is very likely the case, since /h/, /f/, and /θ/ are all "light" fricative sounds, and each occasionally occurs as a very light pattern that could be mislabled as /#/ (blank).

The alternative case, in which a number of dissimilar patterns are misjudged as one pattern, occurs in the case of /m/, /l/, and /i/ being misjudged as /n/. It appears, however, that this pattern of confusions occurs only as the result of simultaneous occurrence of three somewhat unrelated patterns of confusion, i.e., the mutual confusion between /m/ and /n/, the one way confusion between /i/ and the set [/m/, /n/] and the confusion referred to above in which /l/ was sometimes mislabled as one of a number of alternatives. Indeed, it is difficult to think of any one confusion problem that would cause a number of dissimilar pattern types to be mislabled as one category and no such example was found in this study.

Many of the above observations concerning the nature of the confusions indicated by a given submatrix could be used to formulate rules or guidelines for the production of programmed materials to teach the confusions indicated. An initial step in this direction was made at the end of this study. This effort is discussed in a later section (see, Revision.).

5. Programmed Materials

In order to teach discrimination of real and abstract speech patterns self-instructional matching to sample programs were developed. The terminal or criterion behaviors for such programs were: the student should be able to discriminate among the nineteen pattern classes (phonemes) used in this study and further, he should be able to generalize within a class so that he includes all instances which are appropriate to that class. An attempt was made at the outset to identify and make explicit all of the cues which might be relevant to recognition. The analysis involved two major assumptions: first that the subject matter was amenable to a hierarchical arrangement allowing progression from gross to fine discriminations; secondly, that each of the discriminations could be treated as binary. Utilizing these assumptions we partitioned the patterns into two classes on the basis of one distinctive feature. Further sub-classes were created by examining e each of the two classes for another distinctive feature which would allow further subdivision, and so on. The final logic tree from which the program was developed (see Figure 5) was a synthesis of several such trees. It should be stressed that during this phase we developed a number of logic trees each one having face validity and each having certain limitations. One could find in each (including the one in Figure 5) concepts that were not represented as nodes in the tree, nodes that defied concise, accurate verbal descriptions and nodes that represented questionable discriminations. The

A set of rules, or constraints, was added to the basic logic tree. The number of frames in each set was dictated by the number of different phonemes in the set. When two phonemes were being contrasted, for example, each appeared twice as the sample stimulus. Since duplicates of each individual pattern had been photographically reproduced a second rule prohibited the use of the same pattern twice in the same frame.

The tree and the rules produced the design document appearing in Table VI.[1] The actual program was produced simply by assigning particular patterns from the pattern file to each frame. This was done exclusively on the basis of phonetic label without human judgment of the "goodness" of the example.

This simulation of a mechanical procedure was not accidental. It was an attempt to simulate the implementation of the design document by a machine. Were a computer given the logic tree and the population of stimuli accumulated in the file, it would have made selections in the same way we did. One would be naive to assume that the program so generated was optimal or even that it would prove to be effective. The program was developed logically without reference to the pretest data. (A later revision of the program, as will be seen, benefitted not only from these data but also from the error data supplied by this program.) The development procedure is stressed since the procedures used in the development of this program are representative of those engaged in by programmers and other designers of instructional materials: an analysis of terminal behaviors, followed by an intuitive generation of concepts to be taught and the sequence in which they are to be taught. The frames consisted of a sample and two alternatives, one of which matched the sample, the other from the contrasting group. The sample (called X in the tables and on the frames) was mounted in the center of a sheet of black construction paper with the alternatives (A and B) placed on either side. The frames were numbered and identified as to set. Answers appeared on the back of the frames. Each set of frames was intended to teach a single concept.

---

[1] The verbalizations of contrasts in this Table are only crude attempts at stating the discriminations.

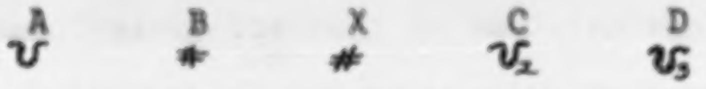
One program contained real speech patterns and the other abstract patterns. The two programs were isomorphic, i.e., a particular real pattern used in the first program appeared in abstracted form in the same place in the second program.

A third program was developed using abstract stimuli and containing four alternative stimuli. An attempt was made to include as many different negative examples (non-matching alternatives) as possible in a single frame. (Figures 6, 7, and 8 present sample frames from the three programs.) A typical frame in this program might be:

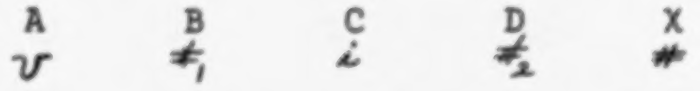
A	B	X	C	D
ʋ	#	#	m	s

Number of alternatives seems to be an interesting variable in the matching-to-sample procedure. It has already been shown to be an important variable in a variety of other experimental procedures (e.g., Riley, 1952). One might propose that this factor should influence discrimination learning for a number of reasons. For example, greater exposure to negative examples, as in the case of numerous alternatives, provides the occasion for more extinction to take place (either implicit or overt). If extinction is indeed "the hallmark of discrimination" (Keller and Schoenfeld, 1950) then with increased extinction finer discrimination should result, or, at least, result more quickly. Yet one study designed to investigate number of alternatives, specifically in the matching-to-sample context (Geis, 1965) failed to show any effect.

Other programming strategies were considered for the four-alternative program. Negative alternatives might, for example, have included only variants of a single class. The above frame might be redesigned as follows:



Another program might include a varying number of positive and negative instances. Instructions to *S* would be: find and indicate which of (instead of which *one* of) the alternatives match the sample. Such a frame might be this one:



Such variations on the basic matching program would, we believe, lead to different searching behavior on *S*'s part and, quite probably, would produce different terminal behaviors than the program we used. Since time was limited, only the first-mentioned four-alternative-program was produced and tested.

All programs consisted of 120 frames. Instructions to the subjects can be found in Appendix A, Exhibit 2. *S* in each case used a program of the same stimulus mode as his pretest (i.e., REAL *S*s used the REAL program). There were five groups. N=3 in each group.

The REAL FORWARD (Group RF) and the ABSTRACT FORWARD (Group AF) groups went through the programs in the sequence indicated in Table VI. The REAL BACKWARD (Group RB) and the ABSTRACT BACKWARD (Group AB) groups did the programs in reverse order (i.e., starting with Set 18 and working to Set 1). ABSTRACT 4 ALTERNATIVE *S*s (Group A4ALT) went through the program forward. (See Table VII for 4ALT Program)

The backward groups were included to test whether or not the sequencing of the program had been successful. If the sequence were optimal or if sequencing of this particular program had any effect at all, the backward groups would show higher frame errors.

Data obtained included; time taken to finish the program, number of errors committed and statements of what concept was being taught in each set. (These reports were of particular aid in revising the program.) Ss' time scores ranged from 20 minutes to two hours with an average of one hour taken to complete the program.

Table VIII. The table below shows total errors per subject and per group.

Group	S1	S2	S3	Total
RB	23	21	21	65
RF	16	19	21	56
A4ALT	17	23	15	55
AB	14	19	15	48
AF	9	21	13	43

Variability among Ss was small; with one exception (S2 in Group AF) within any group Ss showed similar error scores. The differences between groups is not particularly dramatic either. It is surprising to discover that Groups RB and AB showed only slightly poorer performance than the forward groups. Apparently, the sequencing of the program, carefully and logically worked out, had little effect on the error rate. It is also interesting to observe that the abstract programs, while slightly superior to the real programs, are not markedly different from them in terms of error scores. Increasing the number of negative instances in the frame (Group A4ALT) produces somewhat more errors than the two-alternative abstract program; but the Ss in A4ALT still perform as well as the Ss using the real programs. A discussion of the final criterion, the post program test scores, will appear later.

A frame by frame analysis of the program can be found in Table IX. Casual observation indicates that a few sets and even a few frames are primarily responsible for errors. Figure 9 shows this clearly. There, cumulative error is shown as a function of successively ranked frames (i.e., the highest error frame is ranked #1 and produced 13 errors, etc.). Less than 20% of the frames accounted for more than 50% of the total error in the program. Errors concentrated in a few frames means that Ss "agreed" to a large extent on the error frames. Thus, they not only showed small variability in the total number of errors on the program; they also tended err on the same frames.

It will be recalled that each set of frames was designed to produce a single discrimination, concept or rule. To determine which concepts were not being successfully taught sets were analyzed for error rate. The number of Ss was multiplied by the number of frames in the set to produce the denominator of a fraction in which the number of errors was the numerator. Sets with error rates well above 10% were: 3, 4, 8, 10, 15, 16, 17. Sets 8 and 10 represent difficult discriminations — discriminations that are marginal even for an expert due to the ambiguity of stimuli. Sets 15, 16, and 17 involve second order discriminations, i.e., the S must keep several factors in mind at the same time. A number of the discriminations require judgment based on multiple dimensions. Since the program is designed to examine only one dimension at a time, it is not unexpected to find these discriminations difficult.

A concept may be defined as generalization with a class and discrimination between that class and all others. Ss repeatedly demonstrated throughout the program that discrimination training in particular sets was effective. However, the program failed to produce appropriate generalization. All of the samples in a set were not viewed as members of the same class. Ss indicated this defect both by their error scores and in their descriptions of the concepts being established in each set. A failure to generalize could result from Ss searching for the unlike alternative rather than matching the like pattern to the sample. The programming strategy described earlier, which demands that S search all patterns and pick out as many similar stimuli as possible, is probably one way to produce generalization more effectively.

6. Revision of Program

Part of the goal of this study was to evolve rules or techniques for program design using confusion matrix data and in the process generate a revised teaching program. Time permitted testing the program on only one subject. Since testing of the results in this area was rather limited, it should be realized that the comments below should be taken as suggestive rather than definitive. The object at this stage was mostly to define directions for future work.

Some Problems with Format. One of the variables considered in the initial program format was the number of alternatives. Programs were constructed with two and four alternatives. Our experience in constructing these programs, and in analyzing the results (including informal comments from the subjects) indicated that if the number of alternatives is restricted to one fixed number, the format is bound to be non-optimal for the resolution of some discrimination problems. For example, when attention must be directed to the presence of one particular pattern feature, the AXB format, with suitably chosen examples, might be considered optimal. However, when it is necessary to call attention to the range over which a certain variable can vary with a pattern type, it is highly desirable to have a format in which more than two or three patterns can be considered by the subject at one time.

The number of samples in each frame could be increased, or varied from frame to frame; so, too, could the number of alternatives. The size of the population from which samples and alternatives were drawn could be much increased. In the present study, for example, varying the signal level in generating patterns would have enormously increase the size of the stimulus file.

The task itself could be changed while preserving the basic discrimination frame previously worked with. For example, as has been suggested earlier, the subject might be required to detect those alternatives that match (or do not match) a sample from among matching stimuli, the number of which varies on each frame; or he could be required to always find the non-matches, etc.

Two major observations evolved by the time the matrices had been developed and the first programs tested:[1]

(1) The constraints that had been placed on the teaching material--the rigid adherence to the two alternative discrimination frame--may have been necessary for early research but obviously were too broad and, in a sense, too simple, to result in optimal teaching materials, especially when many different kinds of discriminations were involved. The program format, it was concluded, should be made more flexible.

(2) The design of the original program suffered from a serious basic deficiency. The emphasis had been in producing a clear discrimination between stimuli. To a large extent the program was successful in doing that. But as was pointed out in the Introduction, a concept involves generalization as well as discrimination. The program failed to produce responding on the basis of classes or sets. The other side of the coin--generalization within classes--had to be considered at the very outset in redesigning the programmed materials.

It was decided to make an exploratory effort towards constructing a program with a comparatively flexible format. This format would allow an AXB type of presentation, but also would allow the presentation of more than two pattern types at one time, or the presentation of a large number of patterns from one type.

New Experimental Procedures. In the new procedure, using only the REAL stimuli, the "alternative" stimulus, or "unknown", was mounted under a black mask on a metal strip covered with black cloth tape. The construction was similar to the canonical forms except that the metal strips were slightly smaller. The unknowns were presented one at a time to the student within a sequence of "sets" of varying length (up to 15 unknowns/set). For each set a predetermined group of the canonical forms (those used in obtaining the confusion matrix) were arranged on magnetized strips in front of the student, either singly or in groups. The group was placed so that there was room by the side of each canonical form or group of canonical forms for a large number of other forms--the unknowns. Each set of unknowns was designed to resolve one confusion problem, as indicated by the confusion matrix, the problems having been isolated and examined at length as described previously (i.e., see "Pretest").

---

[1] We express our thanks to Dale Brethower of the University of Michigan for his stimulating comments and suggestions at this point in the study.

In each set the student, given the unknowns in a set order, had to match each unknown to one of several of the canonical forms in front of him. In the first set, for example, S̲ was given a set of 9 forms (3 Ɛ and 6 I) with the canonical forms I and Ɛ on the strip before him. The subject was informed if his answer was right or wrong; if wrong, he was allowed to guess again if more than one alternative remained. When the correct answer was finally obtained the unknown was placed alongside the form or forms with which it matched. In this way the student had available for his next matching all the forms previously matched in that set as well as the canonical forms.

Program Development. Since the program was to be derived as algorithmically as possible from the confusion matrix, it was necessary to outline a set of program construction rules. The following is a suggestive (and very incomplete) statement of the rules used:

Each set will be one of the following two types:

> Set type 1 - A vs. B. (or A vs. B vs. C, etc.). Subject must decide is $x = A$ or is $X = B$?

Set type 2 - A vs. (B,C). (or A vs. (B,C,D), etc.) Subject must decide is $x = A$ or is $x = (B \text{ or } C)$?

Starting from the upper left-hand corner of the confusion matrix, for each pattern of confusions (submatrix) one "set" of teaching materials was constructed. The total number of unknowns in any set was roughly proportional to the number of confusions in the submatrix. In addition, the proportion of each pattern type within the set was roughly proportional to the proportion of times that particular type was misjudged (that type was "sent" and another type "received"). The provision was that there should always be some minimum proportion used for each pattern type involved.

One-way confusions were to be considered after all those mutual confusions were considered which involved the pattern types associated with the one-way confusion.

The rules for determining the teaching set type to be used for each confusion type will not be stated explicitly, but can be generally inferred from the final program shown in Table XI. The program had 19 sets in which a total of 161 unknowns (frames) were presented to the student.

Results. Though it was considered beyond the scope of this study to test this last program extensively, one S was run in order to obtain a preliminary estimate of the validity of this approach and the directions for future research. A rather informal procedure was used in which S was encouraged to verbalize comments or criticisms at any time. Matching responses were also verbal and were recorded by the experimenter.

The time required to complete the program was one hour and 35 minutes. The percentage of errors in the program was approximately 20%. Though this was as high as or higher than the error rates for the earlier programs, it must be remembered that the present program was constructed of the materials having the highest error rates, and would have few if any "easy" frames.

The student was given no pretest but did take the standard posttest for REAL patterns. S's posttest score was 71% correct. Referring to Table IV, it may be seen that this score was somewhat better than was obtained by the students going through the earlier AXB type program, although the score was apparently still not as high as was attained by the control group given feedback. Naturally, in interpreting these results one must keep in mind the differences in the length of the training sessions. Also it should be remembered that the subject, in going through the present program, did not have the advantage accorded by taking the pretest.

Within the present program it was of some interest to see if the algorithmic use of the confusion matrix data tended to produce a program in which the set length reflected the number of frames actually needed. The S was asked at the end of each set whether she would have preferred more or fewer examples in that set. Her comments seemed to indicate that the balance in the program between long and short sets was fairly good, with possibly two or three exceptions, but that overall, the program was slightly short. It is of some theoretical interest that a number of comments said, in effect, that the length of a particular set in general was adequate, but that a few more examples of a certain subset of one pattern would have been desirable. (Some examples of subsets of a pattern might be the "lighter" samples of /l/, or those samples of /r/ that occurred in the initial position in the original spoken word.) Since the confusion matrix entries are averaged over pattern types, any error data concerning a subset of a pattern type is not obtainable from the matrix. This means that sophisticated automated programming techniques (such as would be capable of distinguishing between a lighter /l/ and a darker /l/) must be based on a finer breakdown of the confusion data than that provided by the confusion matrix.

7. Control Groups

Twelve control Ss provided a measure of the effect of continued exposure to testing conditions, with and without feedback. For the program groups special teaching material was developed using the test stimuli. The control Ss merely repeated the pretest procedures three additional times. For the feedback group the correct answer (i.e., the number of the matching canonical form) appeared on the back of each sample page. Therefore as S turned to the next sample he could check his previous answer. The no feedback Ss had no answers provided and therefore the control procedure exactly duplicated the pretest procedure.

It is clear from the data found in Table IV and Figure 10 that feedback produced a relatively higher rate of error reduction during the successive runs and resulted in a lower terminal error score than no feedback. This finding is congruent with the results reported in the large body of literature devoted to examining knowledge of results as a variable in learning (e.g., Riley, 1952).

The procedure of no feedback might be thought of not merely as a control but a teaching technique involving only increasing familiarity with the patterns. It will be seen that a small but significant increase in performance can be noted for the no feedback group using the REAL patterns.

The control group with feedback provides a convenient base or reference point from which one might evaluate the performance of more complex teaching methodologies (i.e., programmed instruction). However, it should be kept in mind that the sequential presentation of randomly selected examples may sometimes be a significant teaching methodology in itself.

8. A Comparison of Training Methodologies

As will be recalled from the Experimental Design section, after the intervening treatment (program or control) all Ss were again run on the confusion matrix procedure. For any S this posttest exposed the same stimuli in the same order as in the pretest. Table X compares the percent errors for the various groups before and after treatments. Earlier it had been observed that the ABSTRACT stimuli consistently produced error scores lower than the REAL. In general the same is true for the posttest.

Consider now only the Control groups. It is clear from the table that the Feedback groups are superior to those which had no feedback. This finding is not surprising since it has already been seen that the Feedback groups showed superior performance during the three successive runs which intervened between pretest and posttest (see Control Groups section).

Observation of the results of the Program groups alone reveals that (1) the order of presentation of the program had no more effect on the posttest score than it had had on frame error score within the program: forward and backward groups show almost exactly the same amount of gain; (2) the addition of two more alternatives in the ABSTRACT program not only did not significantly affect error scores in the program, but also did not affect learning as measured by gain between pre- and posttest scores: the 4ALT group performs about like the other ABSTRACT groups on the posttest; (3) the percent gain due to the program treatment is about the same for both the REAL and ABSTRACT groups. (It might be noted parenthetically that the posttest score for the revised program is only 3 percent better than for the AXB program.)

Since the above-mentioned independent variables seemed not to have produced significant differences between groups, it is possible to compare the control groups and the program groups as two units. The average net gains for the control groups and program groups, respectively, are both about 5 percent. Thus, when considering only the posttest scores the programs proved singularly ineffective in producing changes greater than those produced by continued random presentation of the stimuli, especially if the random presentation is accompanied by feedback. This negative result indicates that the terminal effect produced by all the programs used in this study could be produced by much less sophisticated training methodology (as: random presentation with feedback).

So far we have considered only the terminal point reached by each group--the posttest score. An appropriate analysis of the various treatments should include, as well, a comparison of the efficiency of the various methods or treatments. In the present study, for example, both programmed instruction and a random sequential presentation with feedback produce the same or similar gains. Yet the cost, in terms of student time, instructor time, equipment, etc., might be quite different for each treatment. Another factor to be considered is that the posttest score might represent an asymptotic value for one treatment (with no additional improvement possible if the treatment were continued) and a transitional value for another treatment (with an increase in score still attainable through continuation of the treatment). It seems appropriate therefore to study more closely the particular learning curves displayed by each group.

In Figure 11 the percentage of error is shown as a function of the number of frames responded to in which feedback was received. Curves are shown for the AXB programs, the revised program, and the control groups receiving feedback. Since no pretest was taken by the subject taking the revised program, her pretest score is assumed to be the average of all REAL pretest scores. Along the right hand margin are also indicated the posttest scores for the control groups receiving feedback.

The horizontal dashed line indicates the percentage of errors on the revised program. Since the program was specifically constructed from the more difficult discriminations, it may be significant that the percentage of errors was so much smaller than on the posttest.

A number of observations can be made from the data in Figure 11.

The initial rate of learning (in percentage improvement per frame) seems about the same for all teaching methods represented. However in order to compare the rate of learning after, say, 120 frames, it would have been necessary to posttest the feedback control group after 120 frames. Without such a procedure the true effect of the programming methodology on the initial rate of learning must be considered as unknown.

There are a number of 'levels of complexity' that appear to be inherent in the patterns and pattern types used. For brevity we will illustrate this for only the REAL patterns. It appears that approximately 60 percent of the patterns used can be matched to their proper canonical form with no (or little) training; it is only necessary for the student to verify a previously established discrimination. In order to bring out these discriminative responses which are already in the student's repertoire, only a "setting" procedure is called for--one that informs him that the discrimination is appropriate in this context. A second set of discriminations, enabling the student to correctly match up to about 80 percent of the patterns, seemed to be learned within the experiment by many of the subjects. These can apparently be learned as well in a random sequential presentation with feedback as with programmed materials. It can be hypothesized that these discriminations involve simple stimulus dimensions. Still a third set of discriminations not successfully taught by any of the procedures we used may involve multiple dimensions and judgmental responses. We suggest that the student, in order to make such discriminations, must be able not only to "pay attention to" the correct dimensions but must also weight them to form a pattern which is then the basis for the naming, sorting or recognition response. One of the primary goals of the more sophisticated presentation schemes was to enable the student to learn these complex, evasive concepts. For example, consider those concepts that involve the weighting of marginal, conflicting cues. The revised program was specifically designed to produce such conceptual behavior by allowing the student to compare, side-by-side, a sequence of examples of a particular pattern type. It is clear from the figure that the value of the revised program in this regard could not be judged without a considerably longer program. In future work it would be highly desirable to determine the asymptotic behavior of the programming methodology being investigated, such as by using longer sequences with intermittent testing.

9. Transfer

As described above, all Ss received a pretest, a training or control sequence, and a posttest in one set of patterns, REAL or ABSTRACT. After the posttest each subject, with the exception of the one subject taking the revised program, was put through the pretest procedure employing the other set of patterns, REAL or ABSTRACT. This last procedure was termed the Transfer Test. It was felt that the score on the transfer test, when compared to the average score attained by the completely untrained subjects, would give some indication of the transfer of training from one modality to the other. This evaluation of transfer is a crucial prerequisite to the determination of the use of, and the optimal sequencing of, real and abstracted or diagrammatic materials in the training of complex discrimination skills.

It may be seen from Table X that the subjects trained with the ABSTRACT patterns did not achieve higher scores on the transfer test (with REAL patterns) than completely untrained subjects taking this same test as a pretest. In fact, on the average there seemed to be a reduction in score of about 6% that could be attributed to the previous exposure to ABSTRACT patterns.

One might speculate that the small decrease in performance noted might have resulted from the fact that the transfer subjects were "used to" looking at the simplified patterns with only a relatively small number of well-defined features, and as a result, did not take into consideration as many of the numerous, but less well-defined features of the REAL patterns as they would have with no previous exposure to the ABSTRACT patterns. If so, this finding would have a considerable significance. However, the small decrement in performance noted could have conceivably been due to factors related to the experimental conditions. A general "fatigue" factor resulting from the sometimes monotonous testing procedures could conceivably be responsible for some decrease in performance on the transfer test. Clearly more research is needed in this area and fatigue would be one variable that might be better controlled.

It may also be seen from Table X that the subjects trained with REAL patterns achieved scores on the transfer test (with ABSTRACT patterns) that were about the same as the scores of the completely untrained subjects taking this same test as a pretest. This result is of secondary interest in this study, since we are focusing here on the training of skills related to the REAL patterns. As with the previous finding, this result should not be considered definitive because of the limitations of the experimental methodology.

In addition to the factor of fatigue, mentioned above, it could be argued that the transfer test used measured only the initial ability of the subject to make discriminative responses after a training on a very different set of stimuli; it might be more illuminating to measure the rate of learning on the new task (which might show an effect of training on the previous task).

Since it was not demonstrated that high level conceptual behavior was acquired under any of the procedures, it is premature to make any hard and fast conclusions about the use of abstract display when teaching for "real world" recognition.

10. Conclusions and Implications

1) Generating abstractions could be a major task in the design and production of educational materials. In order to arrive at consistent translation of real patterns to the abstract mode a great deal of time was spent in this study in (a) discovering and verbalizing the critical features in the real patterns and (b) deciding upon a set of rules for translating these features into abstract patterns.

2) The confusion matrix data verified a common assumption: students often approach a new subject matter being able to make a majority of the discriminations that will be called for.

3) The design of material for teaching, if it is to teach efficiently, should be evolved with a precise knowledge of students' entering (pretest) behaviors.

4) The confusion matrix model helps define (a) the particular discriminations which require teaching, (b) the degree of difficulty of those discriminations, and (c) the kinds of generalization errors which may be expected as students go through the teaching materials.

5) The study demonstrated the general value of one of the control procedures used. The randomly sequenced feedback procedure proved to be an important experimental probe for discovering the maximum learning that could be expected to occur with minimal pedagogical organization. It provided baseline values for rates of learning and for the terminal point, or asymptote, that is reached with unsystematic presentation.

6) Probably the original programs were less than optimally effective partly because the authors did not base the design upon the confusion matrix data and partly because the program format was rigidly fixed (two alternative matching-to-sample). The revised program, with greater flexibility of format and empirically determined discrimination points (i.e., the contrasts

to be taught) seemed superior to the earlier version. Flexibility in both design and form is recommended as a result of the experiences gained in this study.

7) It is of special significance that details of the revised program were obtained semi-algorithmically from the confusion matrix data. It may be true that programming-by-rule from empirical data may always produce programs that can benefit from review and revision. Nevertheless, it is interesting to note that the program so produced in this study appeared to be superior to the ones generated in the more traditional, logical ways. Programming-by-rule may be especially well suited to teaching discrimination and classification of complex patterns. As principles are derived for obtaining programs from empirical data, it becomes feasible to consider turning over to a machine a large part of the job of frame generation.

8) Since training on one mode (Real or Abstract) did not transfer to the other it may well be that Ss were learning quite different things in each mode. This suggests a strong caution to designers of educational material--transfer cannot be assumed even when, as in this case, every effort was made to produce parallel abstract and real materials.

List of figures

1. Sample spectrogram
2. Experimental design
3. Canonical forms
4. Apparatus
5. Logic tree
6. Photographs of program frames (REAL)
7. Photographs of program frames (ABSTRACT)
8. Photographs of program frames (4ALT)
9. Cumulative program errors
10. Performance of control groups
11. Error scores for Ss with feedback

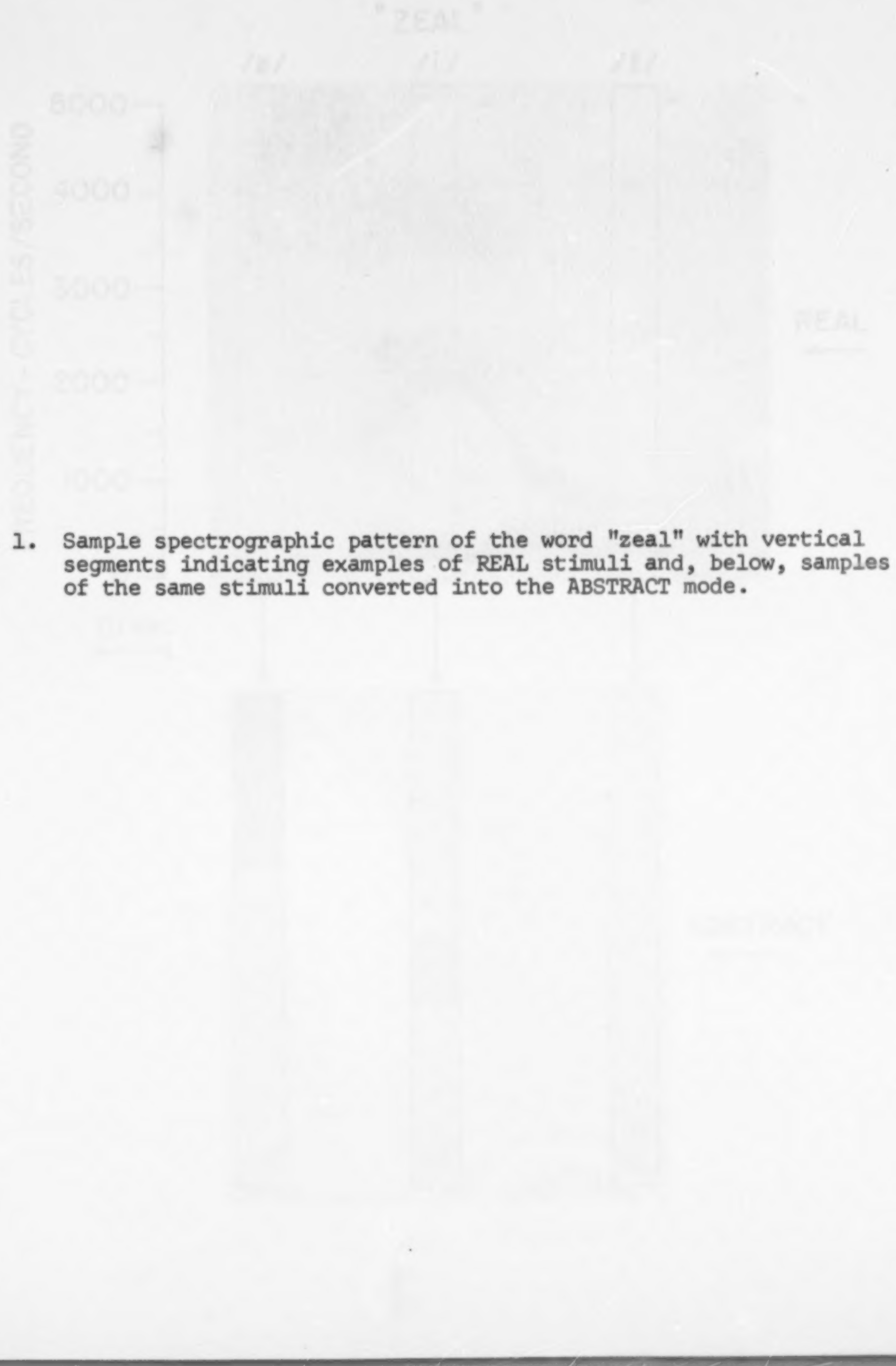


Figure 1. Sample spectrographic pattern of the word "zeal" with vertical segments indicating examples of REAL stimuli and, below, samples of the same stimuli converted into the ABSTRACT mode.

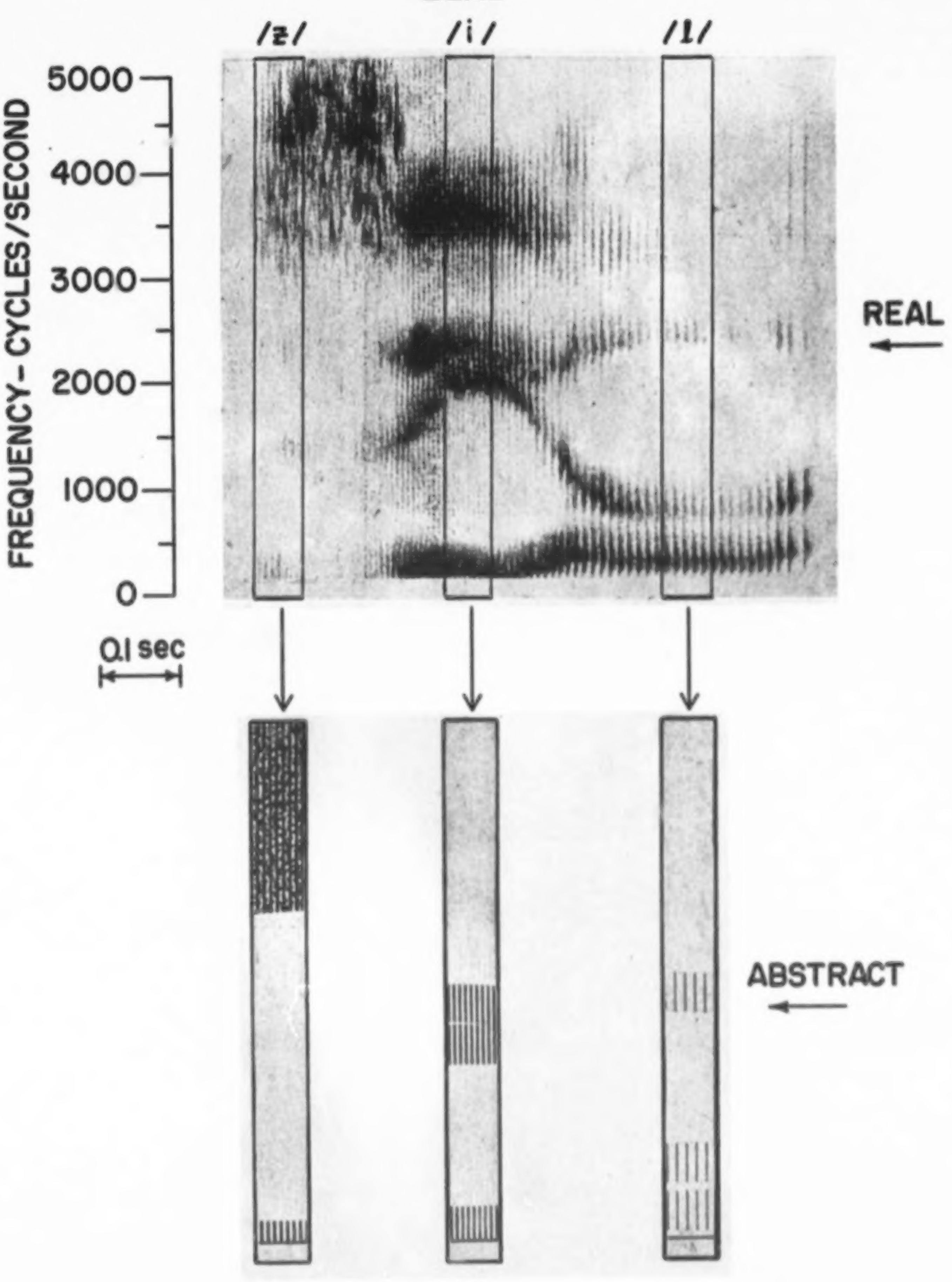

"ZEAL"
/z/
/i/
/l/
5000
4000
3000
2000
1000
0
FREQUENCY- CYCLES/SECOND
0.1 sec
REAL
ABSTRACT

Figure 2. A schematic representation of the experimental design.

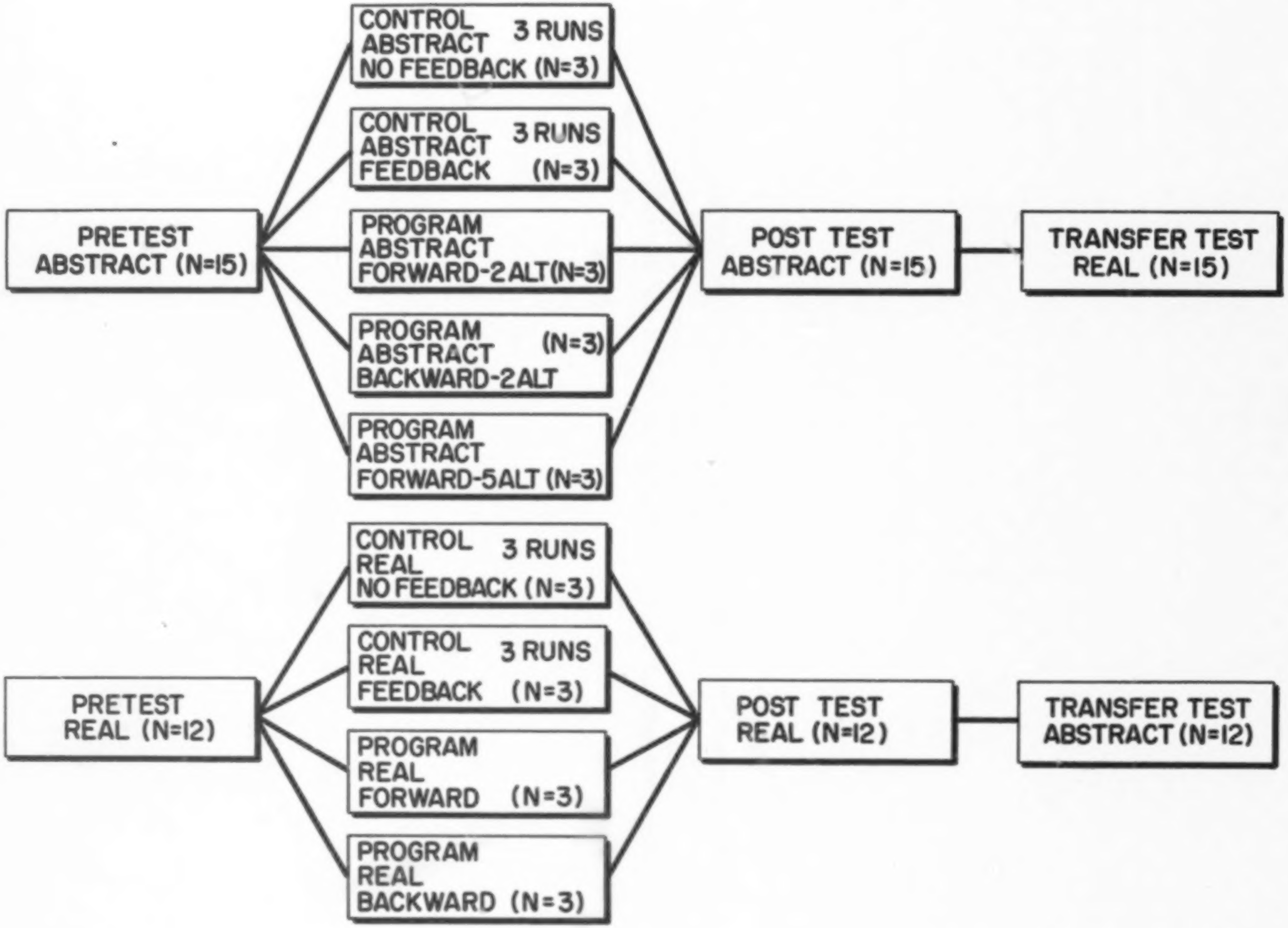
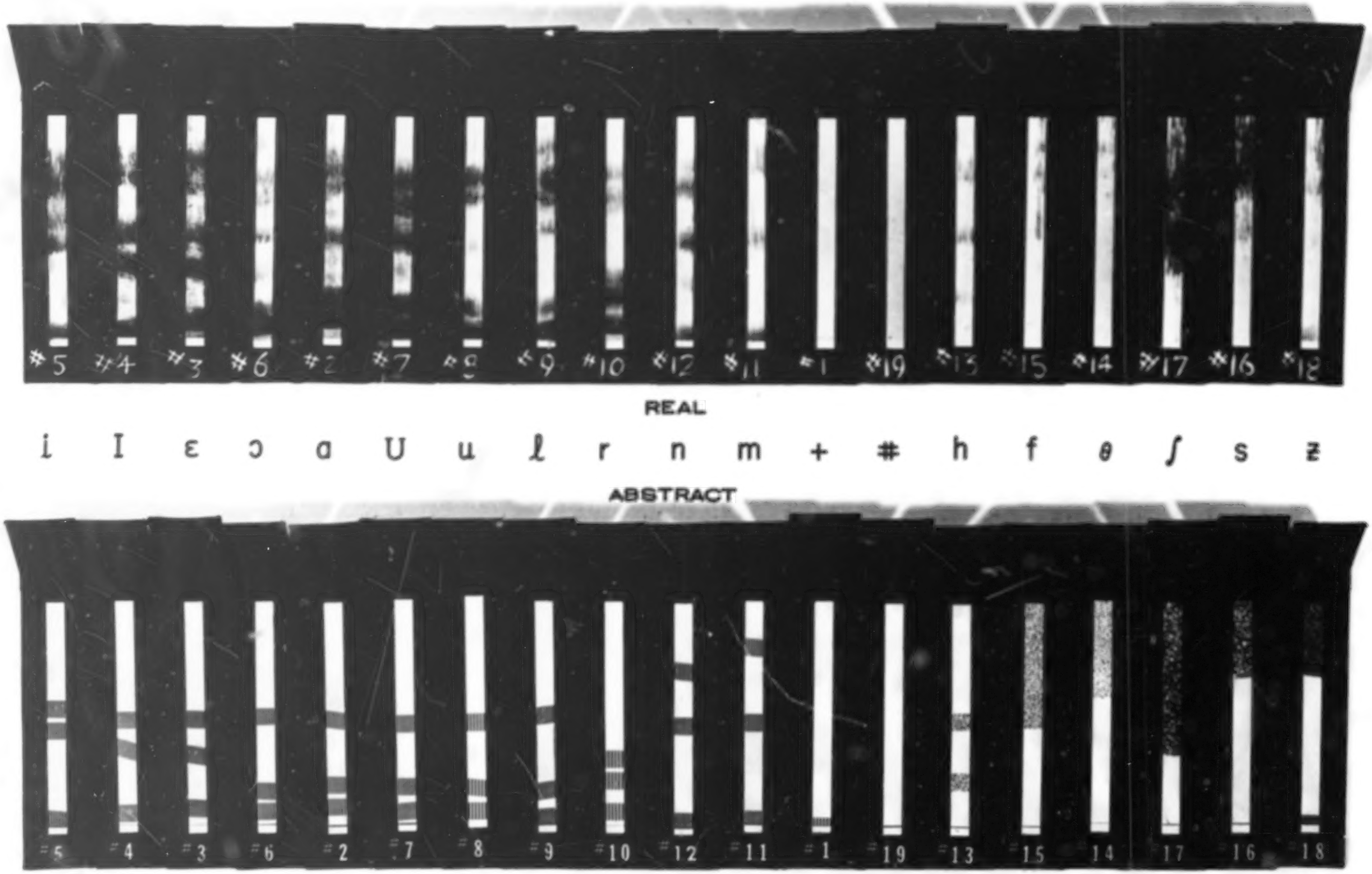

PRETEST
ABSTRACT (N=15)
CONTROL ABSTRACT NO FEEDBACK 3 RUNS (N=3)
CONTROL ABSTRACT FEEDBACK 3 RUNS (N=3)
PROGRAM ABSTRACT FORWARD-2 ALT (N=3)
PROGRAM ABSTRACT BACKWARD-2 ALT (N=3)
PROGRAM ABSTRACT FORWARD-5 ALT (N=3)
POST TEST
ABSTRACT (N=15)
TRANSFER TEST
REAL (N=15)
PRETEST
REAL (N=12)
CONTROL REAL NO FEEDBACK 3 RUNS (N=3)
CONTROL REAL FEEDBACK 3 RUNS (N=3)
PROGRAM REAL FORWARD (N=3)
PROGRAM REAL BACKWARD (N=3)
POST TEST
REAL (N=12)
TRANSFER TEST
ABSTRACT (N=12)

Figure 3. A photograph of the REAL and ABSTRACT canonical forms. The numbers indicate the original order in which the forms were placed on the magnetic bar (with #1 on the left end). The present order of forms corresponds to that in the confusion matrices.


#5 #4 #3 #6 #2 #7 #8 #9 #10 #12 #11 #1 #19 #13 #15 #14 #17 #16 #18
REAL
i I ε ɔ ɑ U u ℓ r n m + # h f θ ʃ s z
ABSTRACT
#5 #4 #3 #6 #2 #7 #8 #9 #10 #12 #11 #1 #19 #13 #15 #14 #17 #16 #18

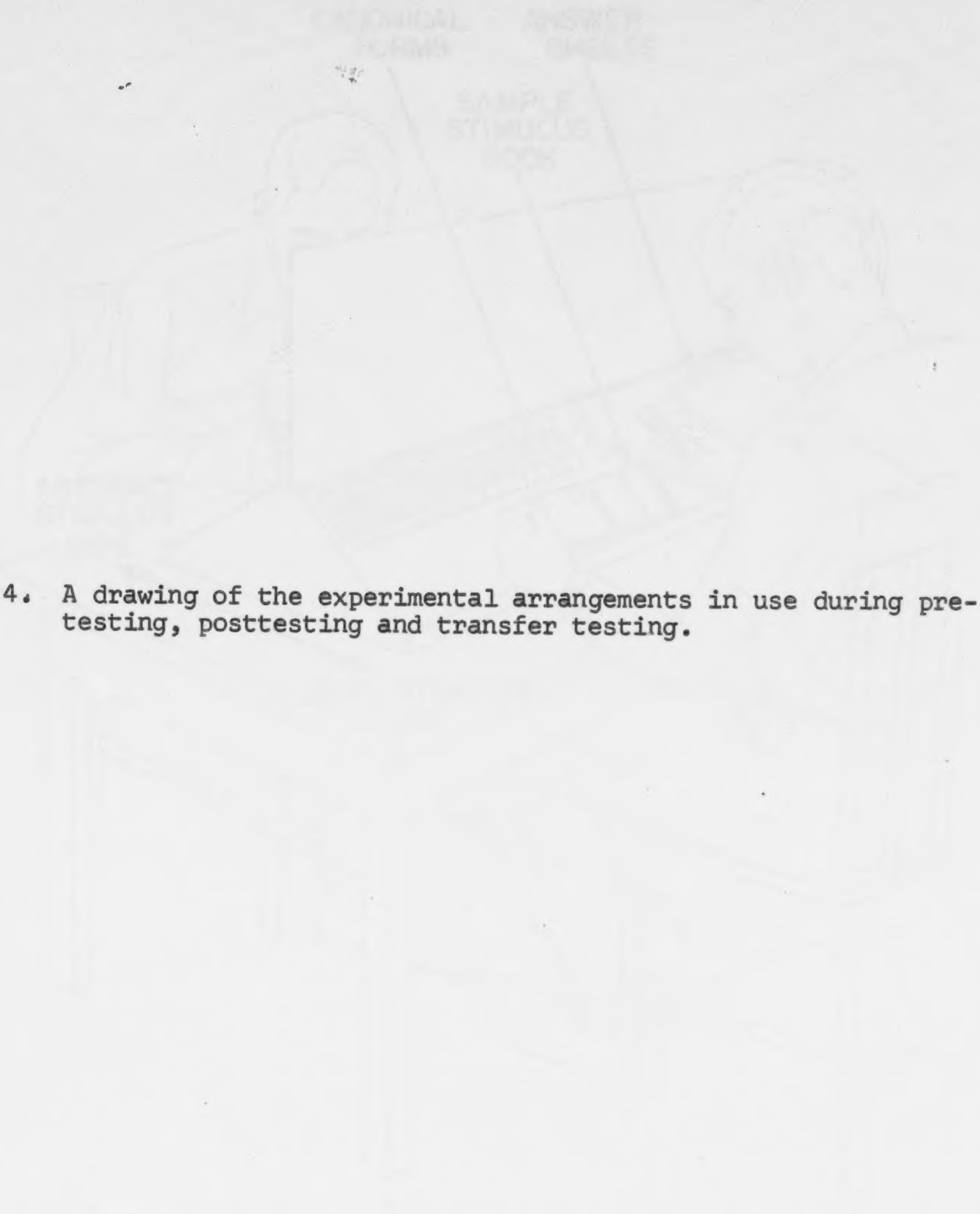

CANONICAL FORMS
ANSWER SHEETS
SAMPLE STIMULUS BOOK

Figure 4. A drawing of the experimental arrangements in use during pretesting, posttesting and transfer testing.

2

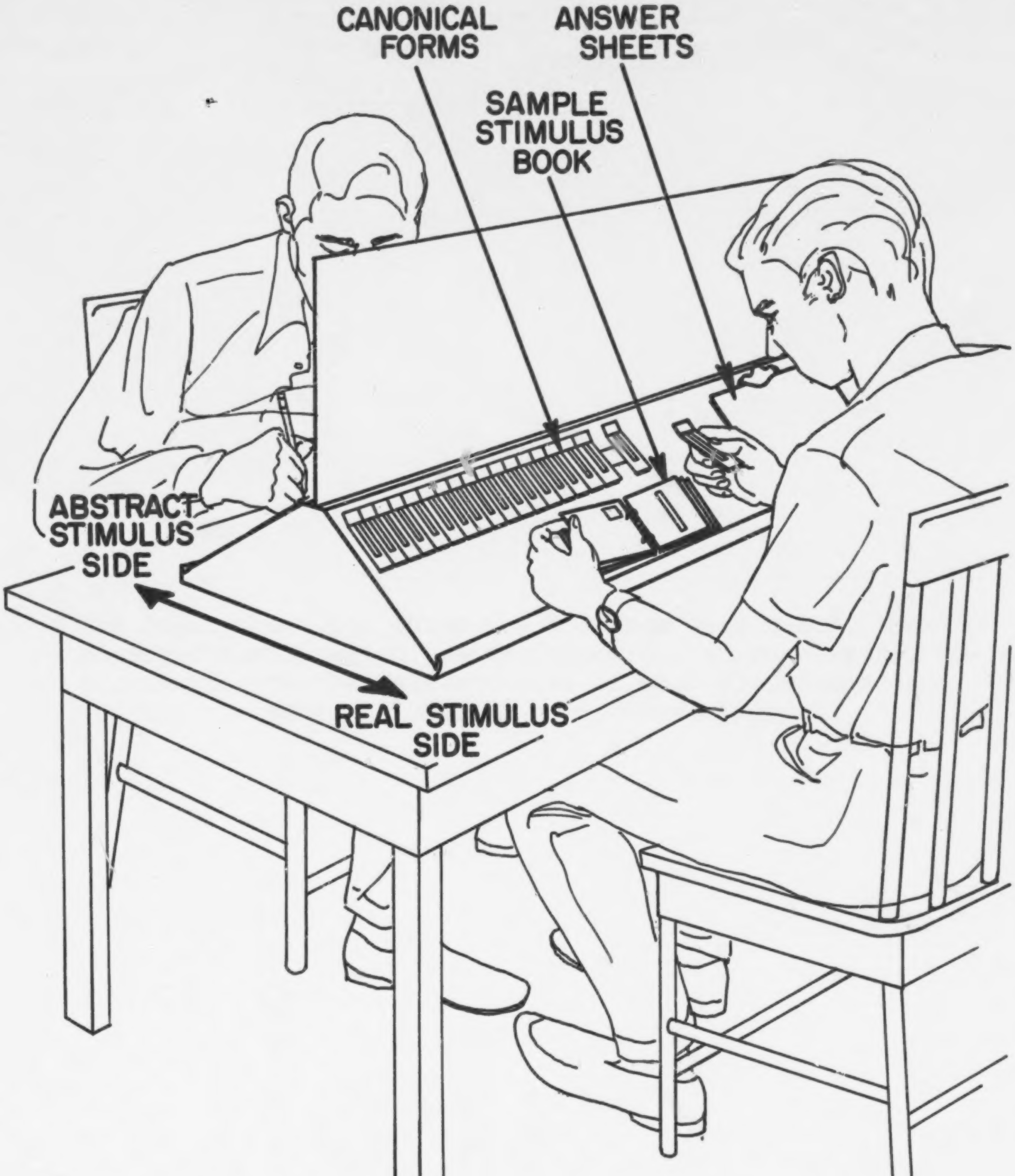
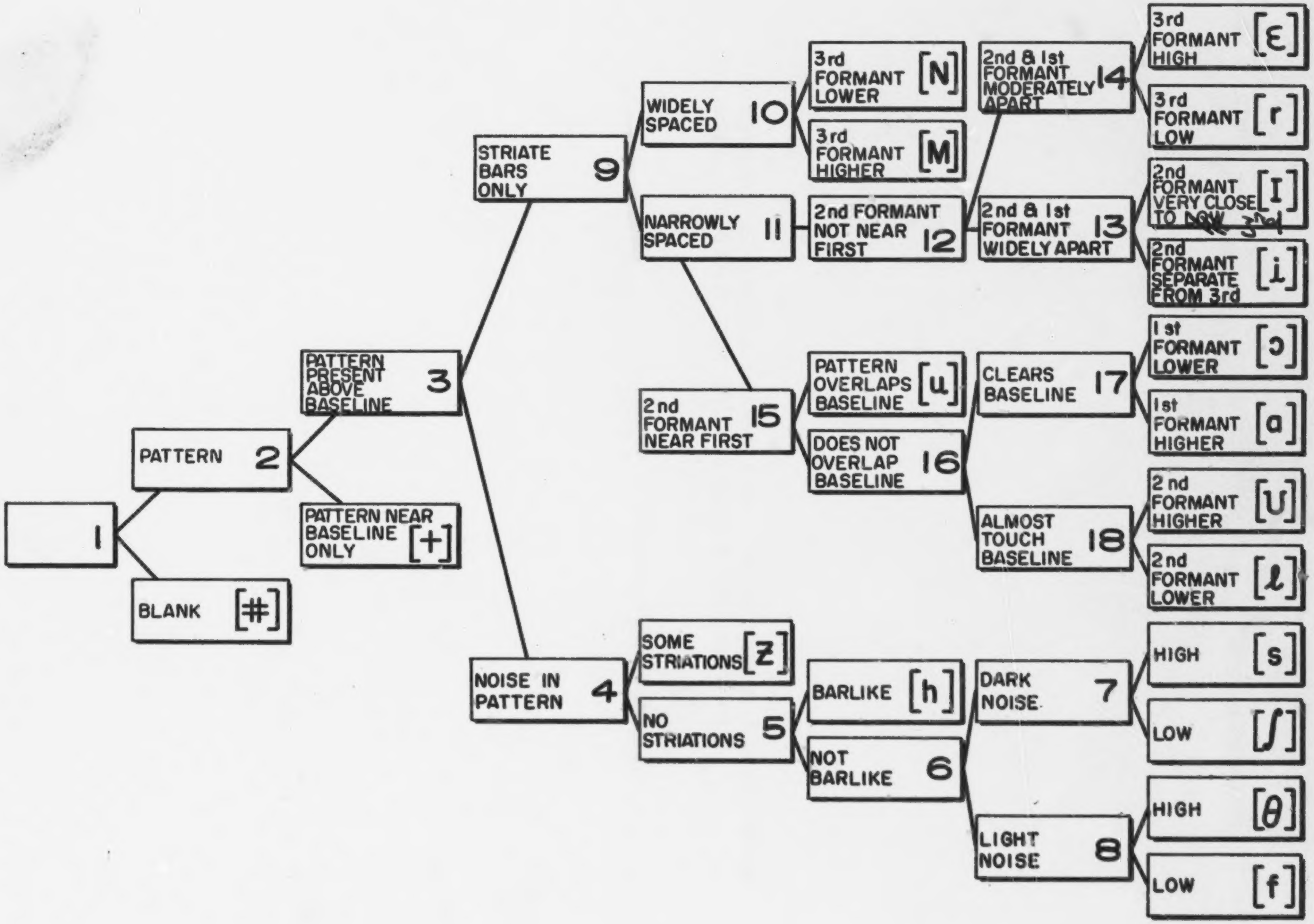
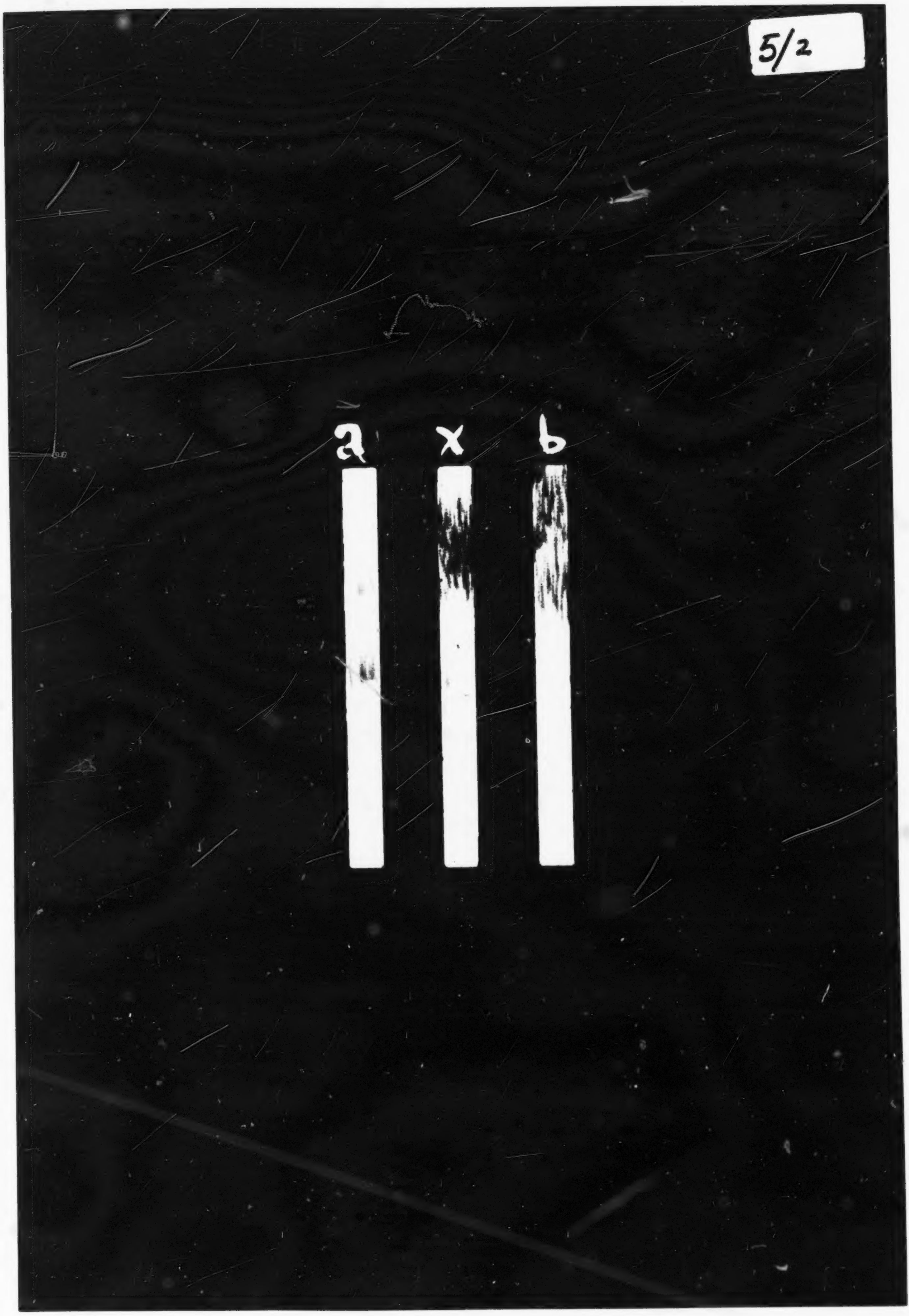

CANONICAL FORMS
ANSWER SHEETS
SAMPLE STIMULUS BOOK
ABSTRACT STIMULUS SIDE
REAL STIMULUS SIDE

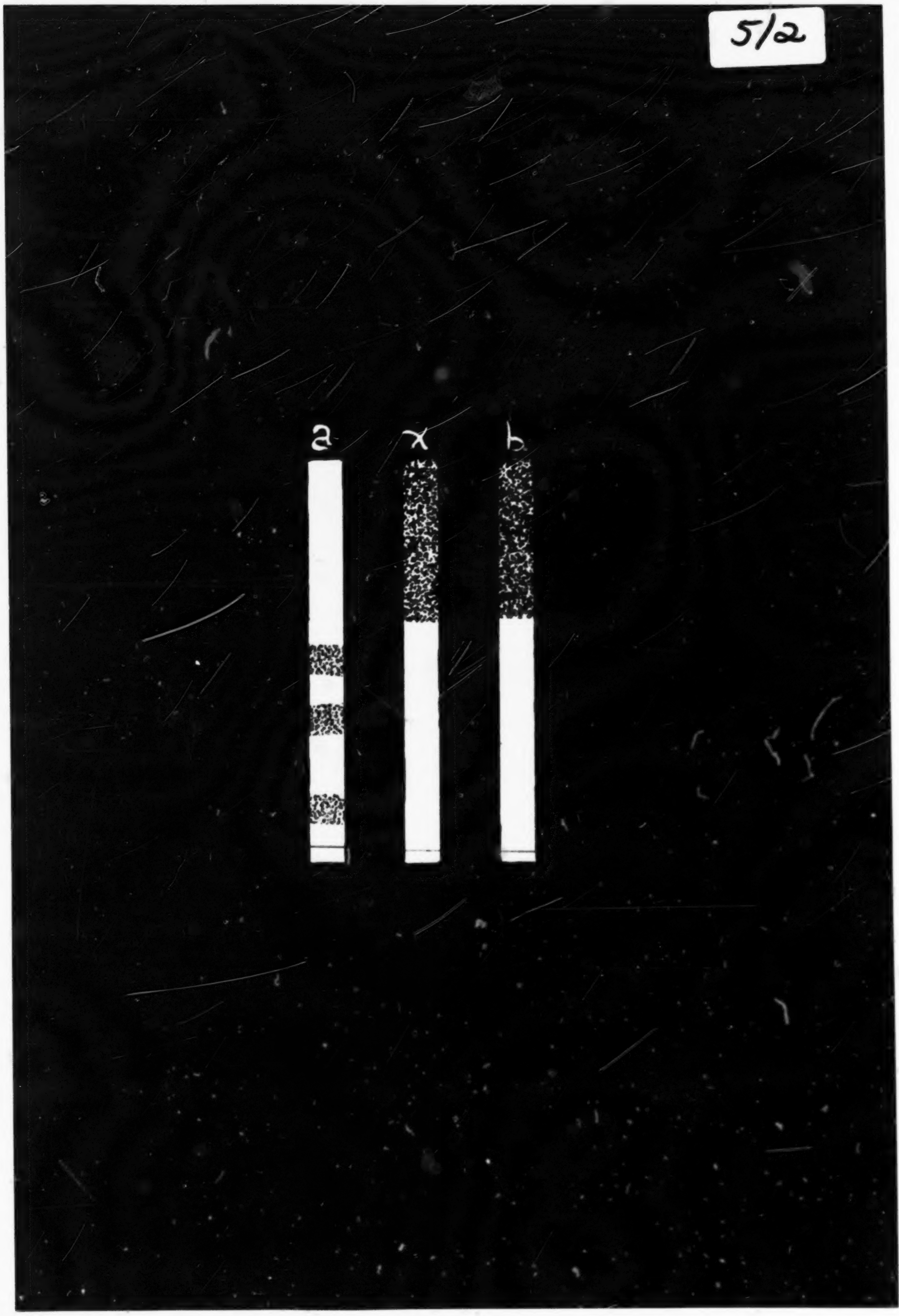
Figure 5. The logic tree upon which the programs were based, showing binary discrimination points (numbers), a rough verbal approximation to the relevant cues of each discrimination, and the phonemes included in each discrimination.


1
PATTERN 2
BLANK [#]
PATTERN PRESENT ABOVE BASELINE 3
PATTERN NEAR BASELINE ONLY [+]
STRIATE BARS ONLY 9
NOISE IN PATTERN 4
WIDELY SPACED 10
NARROWLY SPACED 11
2nd FORMANT NEAR FIRST 15
3rd FORMANT LOWER [N]
3rd FORMANT HIGHER [M]
2nd FORMANT NOT NEAR FIRST 12
2nd & 1st FORMANT MODERATELY APART 14
2nd & 1st FORMANT WIDELY APART 13
3rd FORMANT HIGH [ɛ]
3rd FORMANT LOW [r]
2nd FORMANT VERY CLOSE TO ~~LOW~~ 3rd [I]
2nd FORMANT SEPARATE FROM 3rd [i]
PATTERN OVERLAPS BASELINE [u]
DOES NOT OVERLAP BASELINE 16
CLEARS BASELINE 17
ALMOST TOUCH BASELINE 18
1st FORMANT LOWER [ɔ]
1st FORMANT HIGHER [a]
2nd FORMANT HIGHER [U]
2nd FORMANT LOWER [ℓ]
SOME STRIATIONS [z]
NO STRIATIONS 5
BARLIKE [h]
NOT BARLIKE 6
DARK NOISE 7
LIGHT NOISE 8
HIGH [s]
LOW [ʃ]
HIGH [θ]
LOW [f]

Figs. 6,7,8. Photographs of sample frames from the programs illustrating the REAL, ABSTRACT and 4ALT programs respectively.



5/2


5/2
a
x
b

5/2

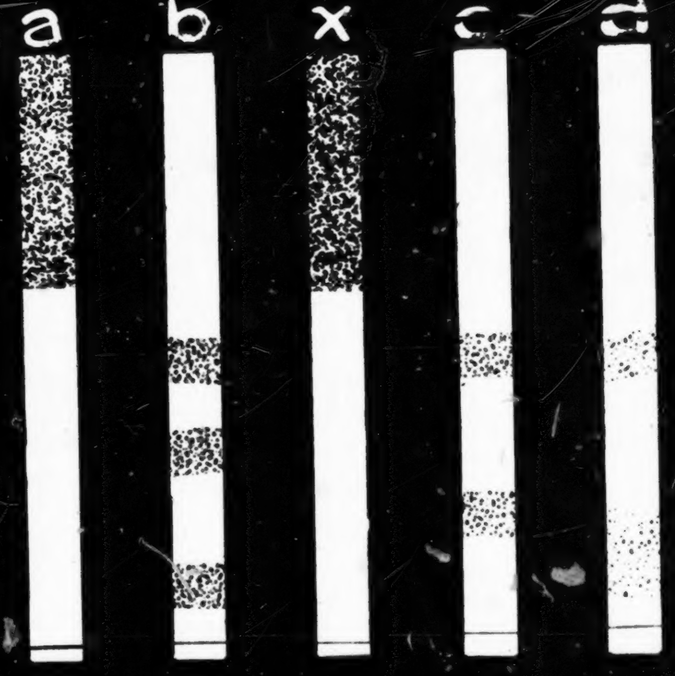


Figure 9. Cumulative errors on the program are plotted as a function of successive frames ranked according to error.

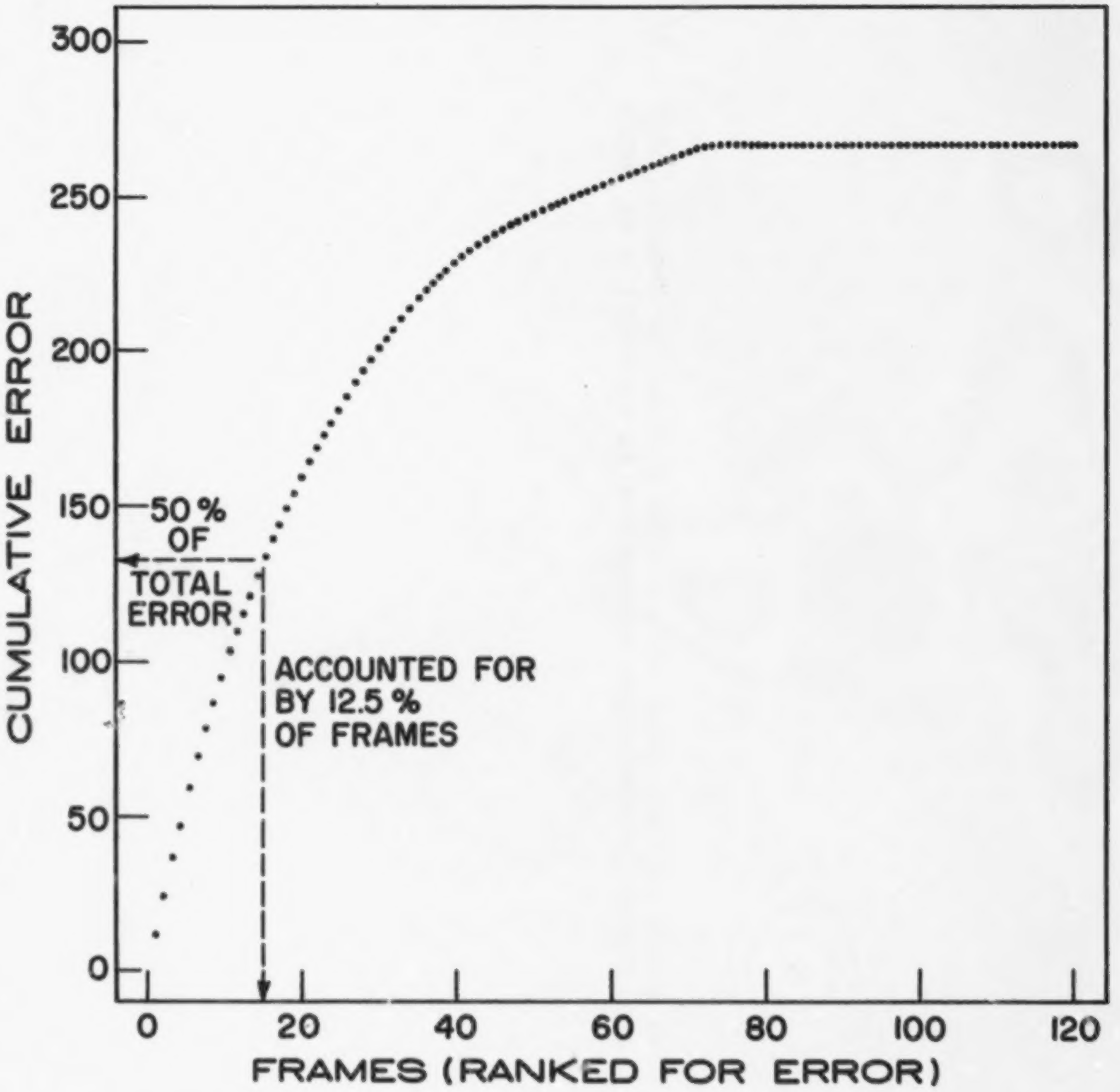

CUMULATIVE ERROR
300
250
200
150
100
50
0
50 % OF TOTAL ERROR
ACCOUNTED FOR BY 12.5 % OF FRAMES
0
20
40
60
80
100
120
FRAMES (RANKED FOR ERROR)

Figure 10. Total number of correct matches for each control group is shown as a function of successive experimental conditions.

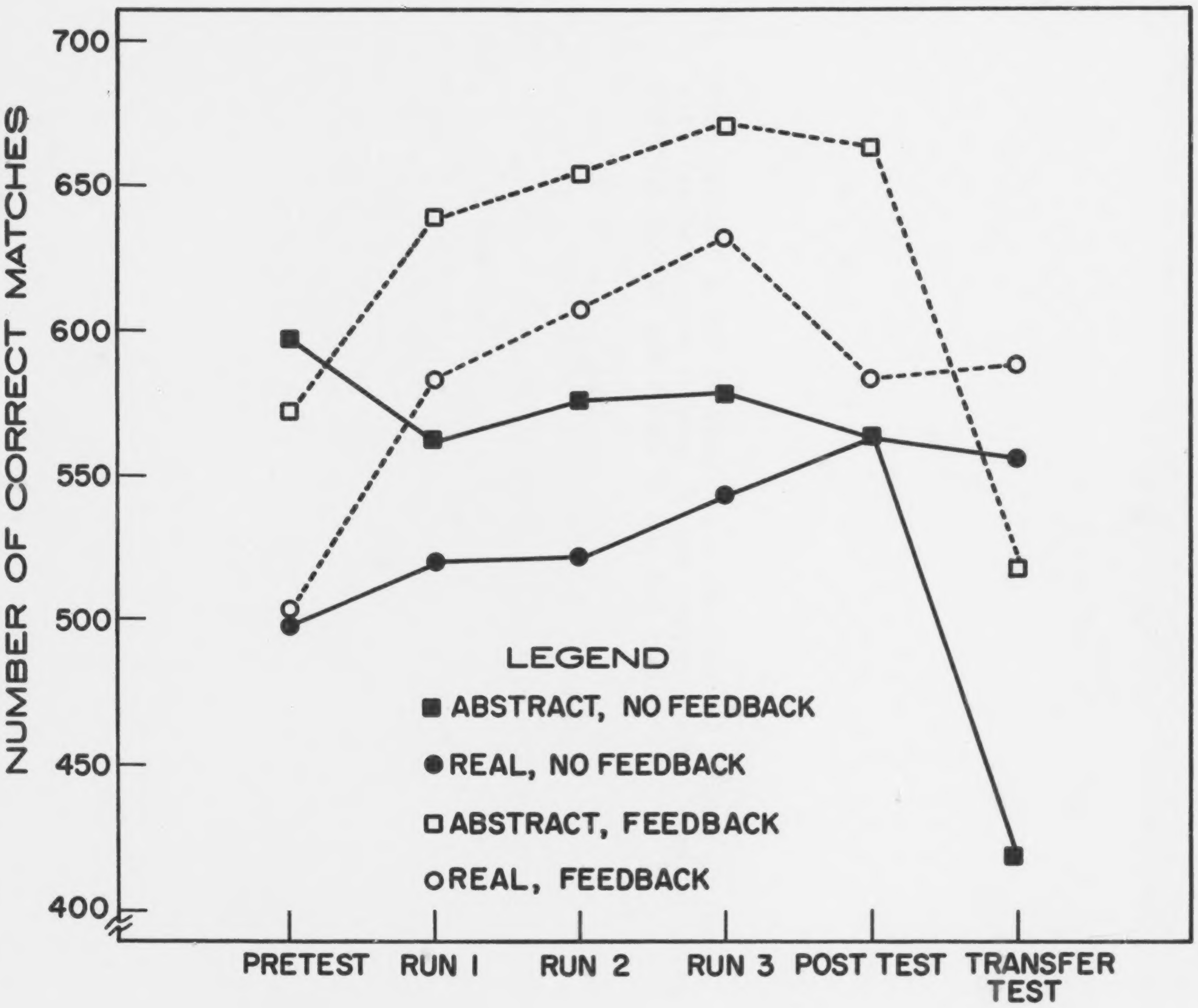
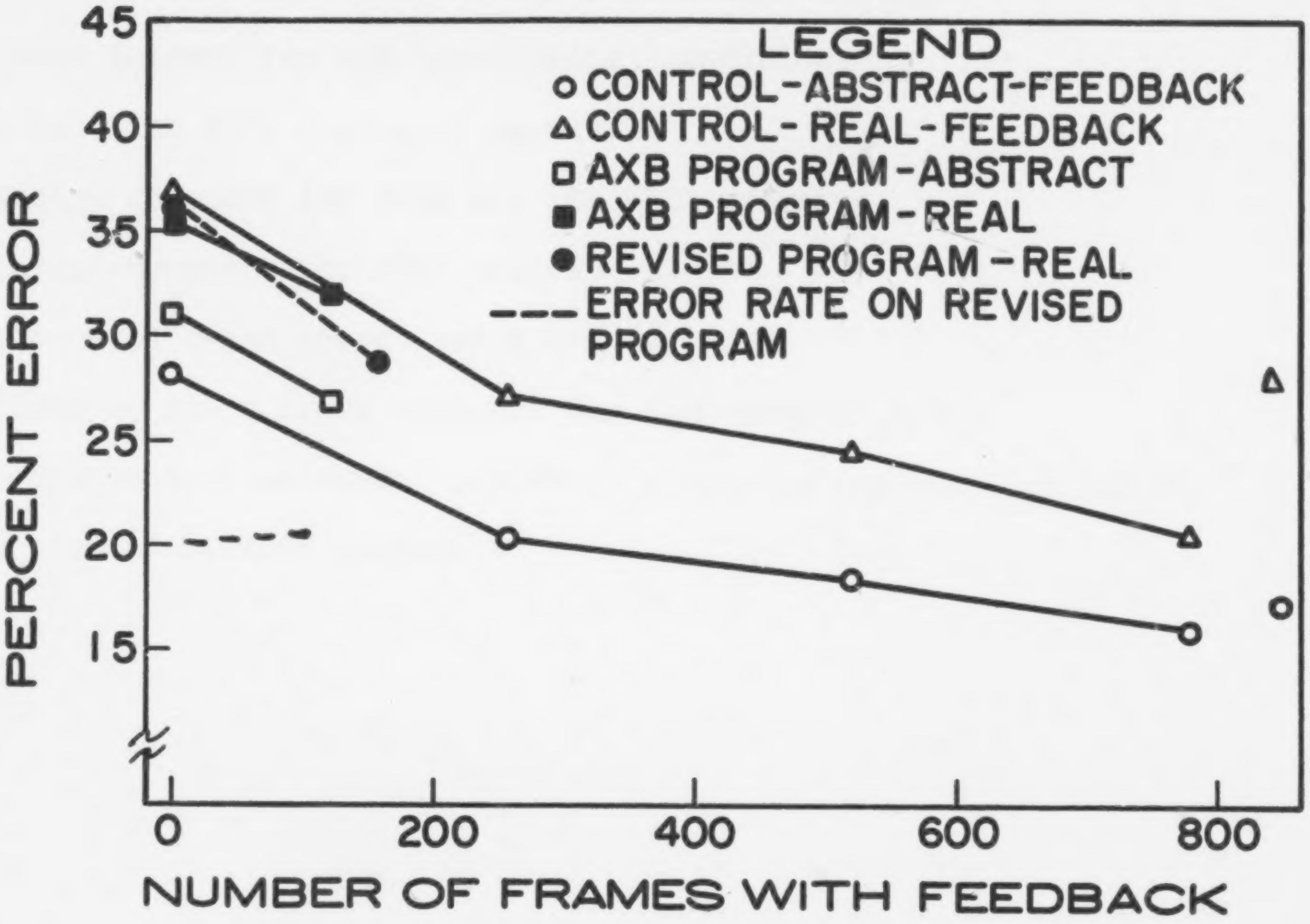

NUMBER OF CORRECT MATCHES
700
650
600
550
500
450
400
PRETEST
RUN 1
RUN 2
RUN 3
POST TEST
TRANSFER TEST
LEGEND
■ ABSTRACT, NO FEEDBACK
● REAL, NO FEEDBACK
□ ABSTRACT, FEEDBACK
○ REAL, FEEDBACK

Figure 11. Mean percent error is shown as a function of number of frames with feedback gone through by the groups shown. Control group posttest scores are shown on the far right.


LEGEND
○ CONTROL-ABSTRACT-FEEDBACK
△ CONTROL-REAL-FEEDBACK
□ AXB PROGRAM-ABSTRACT
■ AXB PROGRAM-REAL
● REVISED PROGRAM-REAL
--- ERROR RATE ON REVISED PROGRAM
PERCENT ERROR
45
40
35
30
25
20
15
0
200
400
600
800
NUMBER OF FRAMES WITH FEEDBACK

List of Tables

1. Phoneme related continuant speech sounds
2. Confusion matrix - pretest using REAL stimuli
3. Confusion matrix - pretest using ABSTRACT stimuli
4. Error summary for all experimental conditions
5. Simplified REAL confusion matrix
6. Design document for REAL and ABSTRACT programs
7. Design document for 4ALT program
8. (In text) frame errors per S and per group
9. Frame by frame error analysis for each program group
10. Percentage of errors on pretests, posttests and transfer tests
11. Design of revised program

TABLE I. A list of those phonemes of English which have a characteristic steady state value that retains its phonemic identity when continued (Dammann, 1965). (Strictly speaking, the voiced hold of a voiced stop and the "silence" states may not fall under this definition, but are included because of their practical significance in instrumental speech analysis.)

Employed in This Study (*)	Symbol	Description
*	/θ/	voiceless dental fricative
*	/f/	voiceless labio-dental fricative
*	/ʃ/	voiceless palatal fricative
*	/s/	voiceless alveolar fricative
*	/z/	voiced alveolar fricative
	/ð/	voiced dental fricative
	/v/	voiced labio-dental fricative
	/ʒ/	voiced palatal fricative
*	/l/	voiced lateral
*	/u/	high back vowel
*	/ʊ/	lower-high back vowel
	/ʌ/	lower-mid unrounded back vowel
*	/ɔ/	lower-mid rounded back vowel
*	/a/	low unrounded back vowel
	/æ/	higher-low front vowel
*	/ɛ/	lower-mid front vowel
*	/I/	lower-high front vowel
*	/i/	high front vowel
*	/r/	retroflexed central semi-vowel
*	/m/	bilabial nasal
*	/n/	alveolar nasal
	/ŋ/	velar nasal
*	/h/	whispered fricative
*	/+/	voiced hold of a voiced stop
*	/#/	interval (silence or non-phonetic noise)

Table II. Confusion matrix for REAL patterns. N=16.
Total errors = 1648 (out of 4238 responses) Pretest.

Received (Canonical form chosen)

Sent (Sample form)

	i	ɪ	ɛ	ɔ	a	ʊ	u	l	r	n	m	+	#	h	f	θ	ʃ	s	z
i	136	14	10	1	1	1	2	10	2	32	29					1			1
ɪ	3	124	62	5	3	7		1		1	1								
ɛ		25	119	30	15	23	2	2	6	1	1								
ɔ			3	83	62	17	27	6	4	2									1
a		3	3	18	160	21	7	8	5	1		1		1					
ʊ				21	49	99	39	9	6		7			6				1	
u	8			1	6	24	124	29	5	16	5	1		1	1	1			2
l	1		1	6	1	2	25	40	42	19	53	8		4		1			21
r				2	1		8	20	124	4	15	34	1			2	1		12
n		2								115	89	1		17					
m	2				1		1	1		44	144	2		11		1	1		16
+												216	4						4
#				1						1		35	202		1			14	
h					1		1			5	5	14	21	83	34	8	9		13
f	1							1			1	2	21	5	142	41	1	6	3
θ		1										3	16	4	26	154	2	9	9
ʃ	2								1								175	30	
s															6	6	4	196	12
z	9						1	4	1	1	3				3	3	3	42	154

Table III. Confusion matrix for ABSTRACT patterns. N = 16.
Total errors = 1375 (out of 4288 responses) Pretest.

Sent (Sample form) \ Received (Canonical form chosen)	i	ɪ	ɛ	ɔ	a	ʊ	u	l	r	n	m	+	#	h	f	θ	ʃ	s	z
i	173	19	3	2	1	1	8	11	7	7	1								7
ɪ	11	85	84	2	3	4	2	9	2	21	2	15							
ɛ	15	18	153	3	18	4	1	5	3	3	1								
ɔ		2	1	81	70	19	26	13	4		1								7
a			14	47	97	35	10	15	4	1									1
ʊ		4	2	42	74	60	24	14	2	1							1		
u	1	3	4	20	42	20	107	19	5	3			3			11	1	1	
l	1	1		44	43	16	26	76	6	4	1								6
r	5	3	1	1	2	1	2	20	179		1							1	8
n		13	5	2	3		3		2	185	11								
m			1		3	1	1	1	3	41	173								
+												209			15				
#												6	218						
h		2	4		2		1	4	2				5	198	1	4		1	
f				2	10	3									149	60			
θ												2	5	2	2	173		40	
ʃ															7		183	1	1
s					1										4	12	30	193	
z				1												1	1		221

Table IV. Total error scores for each group are shown for each test phase

Group	Pre-test	Run 1	Run 2	Run 3	Post-test	Transfer test
Control abs. no feedback	204	238	225	223	238	380
Control abs. feedback	227	161	147	129	136	282
Control real no feedback	303	280	279	257	238	245
Control real feedback	298	217	194	167	218	211
Program abs. forward	263	--	--	--	217	358
Program abs. backward	242	--	--	--	218	363
Program abs. 5ALT	246	--	--	--	211	321
Program real forward	288	--	--	--	258	210
Program real backward	281	--	--	--	256	248
Means						
All abstract groups	237	200	186	176	204	341
All real groups	293	--	--	--	243	229
Program forward groups	276	--	--	--	238	284
Program backward groups	285	--	--	--	229	311
Control feedback groups	263	189	171	148	227	247
Control no feedback groups	254	259	252	240	238	312

TABLE V. A simplified confusion matrix for REAL patterns. N = 16. All entries on the diagonal and those of 16 or below are not shown. Pretest.

Received (Canonical form chosen)

Sent (Sample form)

	i	ɪ	ɛ	ɔ	a	ʊ	u	l	r	n	m	+	#	h	f	θ	ʃ	s	z
i	●									32	29								
ɪ		●	62																
ɛ		25	●	30		23													
ɔ				●	62	21													
a				18	●	17	27												
ʊ				21	49	●	39												
u						24	●	29											
l							25	●	42	19	53								21
r								20	●			34							
n										●	89			17					
m										44	●								
+												●							
#												35	●						
h													21	●	34				
f													21		●	41			
θ															26	●			
ʃ																	●	30	
s																		●	
z																		42	●

Table VI The Real and Abstract Pattern Programs Are Shown in Forward Sequence

Set No.	Frame	A	X	B	Approximate verbal statement of Contrast Discrimination
1	1	n	#	#	No pattern vs. pattern
	2	#	#	ʊ	
	3	#	#	s	
	4	a	#	#	
	5	h	#	#	
	6	#	#	+	
2	1	ʊ	+	+	Voicing vs. all but blank
	2	h	+	+	
	3	i	+	+	
	4	+	+	r	
	5	+	+	I	
	6	+	+	θ	
	7	z	+	+	
3	1	h	h	m	Noise in pattern vs. striate only
	2	z	m	m	
	3	ɛ	ɛ	h	
	4	z	z	u	
	5	ʃ	r	r	
	6	ɔ	ʃ	ʃ	
	7	s	i	i	
	8	r	s	s	
	9	θ	ɔ	ɔ	
	10	i	θ	θ	
	11	f	f	ɛ	
	12	f	u	u	
4	1	z	z	ʃ	Some striation vs. no striation
	2	h	h	z	
	3	ʃ	ʃ	z	
	4	s	s	z	
	5	z	z	s	
	6	ʃ	z	z	
	7	z	θ	θ	
	8	z	f	f	
5	1	ʃ	h	h	Bar-like vs. not bar-like
	2	h	s	s	
	3	θ	h	h	
	4	h	h	s	
	5	h	θ	θ	
	6	h	ʃ	ʃ	
	7	h	f	f	

Set No.	Frame	A	X	B	Approximate verbal statement of Contrast Discrimination
6	1	s	s	θ	Dark vs. light noise
	2	θ	θ	ʃ	
	3	ʃ	f	f	
	4	ʃ	ʃ	θ	
	5	s	θ	θ	
	6	s	s	f	
	7	ʃ	ʃ	f	
	8	ʃ	f	f	
7	1	s	s	ʃ	High vs. low (dark)
	2	ʃ	ʃ	s	
	3	ʃ	s	s	
	4	ʃ	ʃ	s	
8	1	θ	θ	f	High vs. low (light)
	2	f	f	θ	
	3	f	f	θ	
	4	f	θ	θ	
9	1	m	m	r	Widely vs. narrowly spaced
	2	I	m	m	
	3	r	r	n	
	4	m	ɛ	ɛ	
	5	n	n	l	
	6	n	I	I	
	7	ɛ	n	n	
	8	l	l	m	
	9	m	m	i	
10	1	m	m	n	m vs. n, 3rd formant higher vs. lower
	2	m	n	n	
	3	m	m	n	
	4	n	m	m	
	5	n	n	m	
	6	m	n	n	
	7	m	m	n	
11	1	r	r	ʊ	2nd formant not near first vs. 2nd formant near first
	2	ɔ	ɛ	ɛ	
	3	i	a	a	
	4	I	I	u	
	5	ɔ	ɔ	r	
	6	i	i	l	
	7	I	l	l	
	8	a	u	u	
	9	ʊ	ʊ	l	
12	1	i	i	ɛ	2nd formant and 1st formant widely vs. narrowly apart.
	2	r	r	i	
	3	ɛ	ɛ	i	
	4	r	I	I	
	5	I	ɛ	ɛ	
	6	r	i	i	
	7	I	r	r	
	8	I	I	ɛ	

Table VII. The Sequence of Alternative Sources

Set No.	Frame	A	X	B	Approximate verbal statement of Contrast Discrimination
13	1	i	i	I	2nd formant close to vs. separate from 3rd formant
	2	i	I	I	
	3	I	i	i	
	4	i	I	I	
14	1	r	r	ɛ	3rd formant low vs. 3rd formant high
	2	ɛ	r	r	
	3	ɛ	ɛ	r	
	4	r	ɛ	ɛ	
15	1	u	u	ɔ	1st formant touch baseline vs. 1st formant does not touch baseline
	2	ɔ	ɔ	u	
	3	u	u	l	
	4	u	ʊ	ʊ	
	5	u	a	a	
	6	l	l	u	
	7	ʊ	u	u	
16	1	ʊ	ɔ	ɔ	1st formant clears baseline vs. 1st formant almost touches baseline
	2	ʊ	ʊ	a	
	3	l	a	a	
	4	l	l	ɔ	
	5	l	ɔ	ɔ	
	6	ʊ	a	a	
	7	ɔ	ʊ	ʊ	
	8	l	l	a	
17	1	a	ɔ	ɔ	1st formant clears baseline widely vs. 1st formant clears baseline narrowly
	2	ɔ	ɔ	a	
	3	a	a	ɔ	
	4	ɔ	ɔ	a	
18	1	l	l	ʊ	2nd formant higher vs. 2nd formant lower
	2	l	l	ʊ	
	3	ʊ	ʊ	l	
	4	l	ʊ	ʊ	

Table VII. The Abstract 4 Alternative Program

Set No.	Frame	A	B	X	C	D
1	1	n	#	#	s	a
	2	#	ʊ	#	+	n
	3	s	h	#	ʊ	#
	4	a	s	#	#	h
	5	h	n	#	#	ʊ
	6	#	a	#	h	+
2	1	ʊ	+	+	h	θ
	2	h	ʊ	+	I	+
	3	+	i	+	I	ƶ
	4	r	ʊ	+	+	h
	5	I	i	+	r	+
	6	θ	+	+	h	ʊ
	7	+	ƶ	+	r	i
3	1	m	h	h	i	u
	2	ƶ	s	m	ʃ	m
	3	h	ɛ	ɛ	θ	f
	4	u	ɔ	ƶ	ƶ	r
	5	ʃ	ƶ	r	r	h
	6	ɔ	u	ʃ	r	ʃ
	7	s	ƶ	i	h	i
	8	r	s	s	i	ɛ
	9	θ	s	ɔ	ɔ	f
	10	θ	i	θ	ɛ	r
	11	ɛ	i	f	u	f
	12	f	ʃ	u	u	ƶ
4	1	ʃ	h	ƶ	ƶ	s
	2	ƶ	h	h	ƶ	ƶ
	3	ʃ	ƶ	ʃ	ƶ	ƶ
	4	ƶ	ƶ	s	ƶ	s
	5	s	θ	ƶ	ƶ	f
	6	ʃ	ƶ	ƶ	θ	h
	7	ƶ	ƶ	θ	ƶ	θ
	8	f	ƶ	f	ƶ	ƶ
5	1	ʃ	h	h	θ	s
	2	s	h	s	h	h
	3	h	θ	h	ʃ	s
	4	s	θ	h	f	h
	5	h	θ	θ	h	h
	6	h	h	ʃ	h	ʃ
	7	h	h	f	f	h

Table VII (continued)

Set No.	Frame	A	B	X	C	D
6	1	θ	s	s	f	θ
	2	θ	s	θ	s	ʃ
	3	ʃ	ʃ	f	f	s
	4	ʃ	θ	ʃ	f	f
	5	s	θ	θ	s	ʃ
	6	f	s	s	θ	f
	7	f	θ	ʃ	ʃ	θ
	8	ʃ	s	f	s	f
7	1	ʃ	ʃ	s	s	ʃ
	2	ʃ	s	ʃ	s	s
	3	ʃ	ʃ	s	ʃ	s
	4	s	ʃ	ʃ	s	s
8	1	f	f	θ	f	θ
	2	θ	f	f	θ	θ
	3	f	θ	f	θ	θ
	4	f	f	θ	θ	f
9	1	m	r	m	ɛ	I
	2	I	r	m	m	i
	3	n	r	r	m	n
	4	m	ɛ	ɛ	m	l
	5	l	r	n	ɛ	n
	6	n	m	I	I	n
	7	n	ɛ	n	I	i
	8	m	m	l	n	l
	9	l	m	m	r	l
10	1	n	n	m	m	n
	2	m	m	n	m	n
	3	n	m	m	n	n
	4	m	n	m	n	n
	5	m	m	n	n	m
	6	m	m	n	n	m
	7	n	n	m	n	m
11	1	ʊ	ɔ	r	a	r
	2	ɔ	ʊ	ɛ	ɛ	l
	3	i	a	a	I	ɛ
	4	I	u	I	ɔ	a
	5	r	i	ɔ	ɔ	I
	6	l	i	i	u	a
	7	l	I	l	r	i
	8	r	ɛ	u	i	u
	9	ɛ	i	ʊ	r	ʊ

Table VII (continued)

Set No.	Frame	A	B	X	C	D
12	1	ε	i	i	r	ε
	2	i	I	r	r	I
	3	i	i	ε	i	ε
	4	I	r	I	r	ε
	5	I	ε	ε	I	i
	6	i	r	i	r	ε
	7	I	i	r	I	r
	8	I	ε	I	r	ε
13	1	i	I	i	I	I
	2	i	i	I	I	i
	3	I	I	i	I	i
	4	i	I	I	i	i
14	1	ε	ε	r	r	ε
	2	ε	r	r	ε	ε
	3	ε	r	ε	r	r
	4	ε	r	ε	r	r
15	1	u	ɔ	u	ʊ	l
	2	u	ɔ	ɔ	u	u
	3	l	ʊ	u	u	a
	4	u	u	ʊ	u	ʊ
	5	a	u	a	u	u
	6	u	l	l	u	u
	7	ʊ	a	u	ɔ	u
16	1	ʊ	l	ɔ	ɔ	ʊ
	2	a	ʊ	ʊ	a	ɔ
	3	l	ʊ	a	ʊ	a
	4	ɔ	l	l	a	ɔ
	5	l	ɔ	ɔ	ʊ	l
	6	ʊ	l	a	l	a
	7	ɔ	a	ʊ	ʊ	a
	8	a	ɔ	l	ɔ	l
17	1	ɔ	a	ɔ	a	a
	2	a	a	ɔ	ɔ	a
	3	a	ɔ	a	ɔ	ɔ
	4	a	a	ɔ	a	ɔ
18	1	ʊ	ʊ	l	l	ʊ
	2	ʊ	ʊ	l	ʊ	l
	3	l	ʊ	ʊ	l	l
	4	l	l	ʊ	ʊ	l

Table IX Total Errors Committed in Each Group Are Presented for Each Frame and Set

		Program					
Set No.	Frame No.	Real-Forward	Real-Backward	Abstract-2ALT. Forward	Abstract-2ALT Backward	Abstract-4ALT	Raw Total
1	1	--	--	--	--	--	--
	2	--	--	--	--	--	--
	3	--	--	--	--	--	--
	4	--	--	--	--	--	--
	5	--	--	--	--	--	--
	6	--	--	--	--	--	--
	Total	0	0	0	0	0	0
2	1	--	--	--	--	--	--
	2	2	--	--	--	--	2
	3	--	--	--	--	--	--
	4	1	--	--	--	--	1
	5	--	--	--	--	--	--
	6	--	--	--	--	--	--
	7	--	--	--	--	--	--
	Total	3	0	0	0	0	3
3	1	--	--	--	--	--	--
	2	--	--	--	--	--	--
	3	--	--	--	--	--	--
	4	--	1	--	--	--	1
	5	--	--	--	--	--	--
	6	--	--	--	--	--	--
	7	--	--	--	--	--	--
	8	--	--	--	--	--	--
	9	--	--	--	--	--	--
	10	3	3	2	3	--	11
	11	3	3	1	2	--	9
	12	3	3	2	--	--	8
	Total	9	10	5	5	0	29
4	1	3	3	3	3	--	12
	2	2	--	3	3	--	8
	3	--	--	--	--	--	--
	4	3	2	--	--	--	5
	5	1	1	--	1	1	4
	6	--	1	--	--	--	1
	7	--	--	--	--	--	--
	8	3	3	2	3	--	11
	Total	12	10	8	10	1	41

Table IX cont.

		Program					
Set No.	Frame No.	Real-Forward	Real-Backward	Abstract-2ALT Forward	Abstract-2ALT Backward	Abstract 4ALT	Raw Total
5	1	--	1	--	--	--	1
	2	1	--	--	--	--	1
	3	1	--	--	--	--	1
	4	--	--	--	--	--	--
	5	--	2	--	--	--	2
	6	--	--	--	--	--	--
	7	--	--	--	--	--	--
	Total	2	3	0	0	0	5
6	1	--	--	--	--	--	--
	2	--	--	1	2	2	5
	3	--	--	--	--	3	3
	4	--	--	--	--	1	1
	5	--	1	1	--	--	2
	6	1	--	--	--	--	1
	7	--	--	--	--	--	--
	8	--	--	1	--	--	1
	Total	1	1	3	2	6	13
7	1	--	--	--	--	1	1
	2	--	--	--	--	--	--
	3	--	--	--	--	2	2
	4	--	--	--	--	--	--
	Total	0	0	0	0	3	3
8	1	--	3	3	3	3	12
	2	--	3	1	--	--	4
	3	--	--	1	2	--	3
	4	--	--	1	1	2	4
	Total	0	6	6	6	5	23
9	1	--	--	--	--	1	1
	2	--	--	--	--	--	--
	3	1	--	--	--	--	1
	4	--	--	--	--	--	--
	5	--	--	--	--	--	--
	6	--	--	--	--	--	--
	7	--	--	--	--	--	--
	8	--	--	--	--	--	--
	9	--	1	--	--	--	1
	Total	1	1	0	0	1	3
10	1	1	2	--	--	2	5
	2	3	3	--	--	--	6
	3	--	--	1	1	1	3
	4	1	1	--	--	--	2
	5	2	1	--	--	--	3
	6	3	3	3	3	--	12
	7	1	2	1	--	--	4
	Total	11	12	5	4	3	35

Table IX cont.

		Program					
Set No.	Frame No.	Real-Forward	Real-Backward	Abstract-2ALT Forward	Abstract-2ALT Backward	Abstract-4ALT	Raw Total
11	1	3	2	--	--	--	5
	2	1	--	--	--	--	1
	3	1	1	--	--	--	2
	4	--	--	--	--	--	--
	5	--	--	1	--	--	1
	*6	--	--	--	--	--	--
	7	--	--	--	--	--	--
	*8	2	2	--	--	--	4
	*9	--	--	--	--	--	--
	Total	7	5	1	0	0	13
12	1	--	--	--	--	1	1
	2	--	--	--	--	--	--
	3	--	--	--	--	--	--
	4	--	--	--	--	3	3
	5	--	1	--	--	3	4
	6	--	--	--	--	--	--
	7	--	--	--	--	1	1
	8	1	1	1	1	2	6
	Total	1	2	1	1	10	15
13	1	--	--	--	--	1	1
	2	--	--	2	2	2	6
	3	--	--	--	--	--	--
	4	--	--	--	--	1	1
	Total	0	0	2	2	4	8
14	1	--	--	--	--	--	--
	2	--	1	--	--	--	1
	3	--	--	--	--	--	--
	4	--	--	--	--	--	--
	Total	0	1	0	0	0	1
15	1	--	--	--	--	1	1
	2	--	--	--	--	--	0
	3	--	--	--	--	1	1
	4	--	--	1	2	2	5
	5	--	--	--	--	1	1
	6	--	--	2	1	3	6
	7	3	3	2	2	3	13
	Total	3	3	5	5	11	27
16	1	1	2	--	--	--	3
	2	--	1	1	2	1	5
	3	--	--	--	--	1	1
	4	2	2	--	--	--	4
	5	--	--	--	--	2	2
	6	--	1	--	--	--	1
	7	1	2	--	2	3	8
	8	--	--	1	1	--	2
	Total	4	8	2	5	7	26

Table IX cont.

		Program					
Set No.	Frame No.	Real-Forward	Real-Backward	Abstract-2ALT Forward	Abstract-2ALT Backward	Abstract-4ALT	Total
17	1	--	2	2	--	--	4
	2	2	1	1	2	1	7
	3	--	--	--	--	1	1
	4	--	--	--	2	1	3
	Total	2	3	3	4	3	15
18	1	--	--	--	--	--	--
	2	--	--	--	--	--	--
	3	--	--	1	1	--	2
	4	--	--	1	3	1	5
	Total	0	0	2	4	1	7
Sum of All Frames (N=120)		56	65	43	48	55	267

Table X Percentage of erroneous matching responses is shown for each group and for each test.

Group	Pretest	Posttest	Transfer Test
Control Abstract No Feedback	25	30	48
Control Abstract Feedback	28	20	35
Control Real No Feedback	38	30	31
Control Real Feedback	37	27	26
Program Abstract Forward	33	27	45
Program Abstract Backward	31	27	45
Program Abstract 4 Alternative	31	26	40
Program Real Forward	36	32	31
Program Real Backward	36	32	31
Revised Program	--	(29)	--

Table XI The Program Constructed from Real Confusion Matrix

Set	Pattern Types Compared*	Number of Unknowns of Each Type
1	ɛ vs. I	3ɛ, 6 I
2	ɔ vs. a	6ɔ, 2 a
3	a vs. ʊ vs. ɔ	3a, 7ʊ, 5ɔ
4	ɛ vs. (I, ɔ, ʊ)	6ɛ, 1 I, 2ɔ, 1ʊ
5	ʊ vs. u	4ʊ, 2 u
6	u vs. l	3u, 3 l
7	a vs. u	3 a, 1 u
8	l vs. r	4 l, 2 r
9	r vs. (l, +)	4 r, 1 l, 1 +
10	m vs. n	4 m, 7 n
11	i vs. (m, n)	6 i, 2 m, 2 n
12	l vs. (u, r, m, n, z)	12 l, 1 u, 2 r, 5 m, 2 n, 2 z
13	n vs. (m, h)	6 n, 2 m, 1 h
14	# vs. +	4 #, 1 +
15	f vs. θ	4 f, 3 θ
16	h vs. (f, #)	5 h, 2 f, 1 #
17	# vs. (h, f, θ)	3 #, 1 h, 2 f, 2 θ
18	ʃ vs. s	3ʃ, 1 s
19	ƶ vs. s	4 ƶ, 1 s
		Total = 161

*Canonical forms were available for these patterns during presentation of the appropriate unknowns.

References

Bilodeau, E. A. and Bilodeau, I. McD. Motor-skills learning in *Annual Review of Psychology*, Palo Alto: Annual Reviews, Inc., 1961.

Dawkins, P. B. Programed instruction and low altitude aerial observation. HumRRO research report No. 14, U. S. Army Aviation Human Research Unit, Fort Rucker, Alabama, December 1964.

Damman, J. E. Application of adaptive threshold elements to the recognition of acoustic-phonetic states. *J. Acous. Soc. Amer.*, 38, No. 2, August 1965, 213-223.

Geis, G. L. and Spenser, D. An annotated bibliography of matching-to-sample studies. Progress Report No. 1, Center for Research on Language and Language Behavior, University of Michigan, 1965.

Geis, G. L. Matching to sample by retarded children: the effects of delayed response and number of alternatives. Doctoral thesis: University Microfilms, Ann Arbor, 1965.

Heidbreder, E. The attainment of concepts. I. Terminology and method. *J. gen. Psychol.*, 35, 1946, 173-184.

Hively, W. Programming stimuli in matching to sample. *J. exp. anal. Behav.*, *5*, 1962, 279-298.

Hunt, E. B. *Concept Learning: an information processing problem*. John Wiley and Sons. New York: 1962.

Keller, F. S. and Schoenfeld, W. N. *Principles of Psychology*. Appleton-Century Crofts, Inc., New York: 1950.

Lehiste, I. and Peterson, G. E. Vowel amplitude and phonetic stress in American English. *J. Acous. Soc. Amer.*, 31, 1959, 428-435.

Potter, R. K., Kopp, G. A. and Green, H. C. *Visible Speech*. Van Nostrand, New York: 1947.

Riley, D. A. Rote learning as a function of distribution of practice and complexity of situation. *J. exp. Psychol.*, 43, 1952, 88-95.

Rothenberg, M. *Programmed Learning Problem Set*. Edwards Letter Shop, Ann Arbor: 1963.

Uhr, L. and Vossler, C. Recognition of speech by a computer program that was written to simulate a model for human visual recognition. *J. Acous. Soc. Amer.*, 33, 1961, p. 1426.

Weisz, A. Z. and McElroy, L. S. Response and feedback techniques for automated training of visual identification skills. Tech. report: NAVTRADEVCEN 789-3, Port Washington, July 1964.

Appendix A. Exhibit 1 is a copy of instructions given to subjects at the beginning of the pretest. Exhibit 2 is a copy of instructions given to subjects as they began the program.

TO THE SUBJECT:

Your task will be to match one of the forms appearing on the magnetic strip at the top of the board with each of the forms in the book. You are to select and compare among the top forms until you find the one that seems to best match the form you are then looking at in the book. Please follow these instructions:

1) Choose a pattern from among the 19 on the top that seems likely to match the form in the book. You may then take the top pattern(s) down (two at a time if you wish) and place them next to the form in the book for comparison.

2) Do not expect the match to be exact. Select the top form that seems closest to the one you are looking at in the book.

3) Write on the response sheet the number of the pattern from the top row you have chosen as the " best match " next to the number for the item in the book. The appropriate number in the book is that facing the exposed pattern - be sure to locate that number on the response sheet and write your answer next to it. The number for each of the matching forms is on its lower edge (ie. # 1).

4) Before going on to the next page in the book, make a choice even if it is a " guess ". When you write down your choice, put the magnetic strips back in order and turn to the next page in the book. Repeat the matching process for each page in the book.

5) Work rapidly, but carefully. Make a response for each form in the book. When you are done with a book (try to make it about 7 minutes), ask for the next one.

Exhibit 2. Instructions for use of the program

This phase of the experiment consists of matching patterns similar to those you matched in the first three sessions. In the following pages, for each pattern labeled "X" you are to find the "best match". I.e., you are to determine which pattern (A or B) most closely resembles pattern "X".* Do not expect the match to be exact. After you have chosen the "best match", indicate your choice next to the appropriate item number on the answer sheet. The number on the answer sheet should correspond with the number on the upper right-hand corner of the black page. Before going on to the next black page, make a choice, even if it is a guess. The correct answer for each match is on the reverse side of the corresponding black page. Do not look at this answer until you have indicated your choice on the answer sheet.

Each of the 18 sets (a set consists of anywhere between four and twelve black sheets) in this portion of the experiment is designed to illustrate a concept or concepts which, if learned, will aid you in matching patterns. At the end of each set you are to indicate on the reverse side of the answer sheet what you believe the concept(s) for that set is (are). You may look back at previous items if you desire to do so. Do not become discouraged if you are having difficulty in discovering a concept, as many of them are very subtle. However, try to identify the concept(s) for each set, even if you must guess. It is to your advantage to learn as many of the concepts as you can, since your score on a final exam will, in part, determine how much you are paid.

---

*For group A4ALT this sentence read: "I.e., you are to determine which pattern (A, B, C, or D) most closely resembles pattern X.

Appendix B. A copy of a talk given by G. Geis at the 1965 meetings of the National Society for Programmed Instruction. Note: The Figure 2 referred to in the talk is Figure 1 in this report.

Visible Speech: Examples, Discrimination, and Discrimination of Examples

George L. Geis
Center for Research on Learning and Teaching
The University of Michigan

Delivered at National Society for Programmed Instruction Convention
May, 1965

The statement is frequently made that programs are designed to shape new behaviors. I would contend that in many cases this simply is not true. Programs, I think, are most often designed to bring old responses under the control of new stimuli, or to put it another way, are designed to produce new discriminations or concepts using responses familiar to the subject. Most psychologists would agree, at least in reference to practical teaching situations, that discrimination learning, or more broadly concept formation, involves the presentation of positive and negative instances of the concept. Were we teaching the concept bottle we would present examples of bottles and non-bottles, the non-bottles including bottle-like objects such as jars and glasses. With this approach in mind a number of people at The University of Michigan have evolved what we call a discrimination frame, an example of which is shown in Figure 1 on the handout. Typically, such a frame presents three objects or units, one which is identified as an example of the concept, the other as not an example of the concept. The student decides whether the third object or example is a member of the concept class or not. The responses involved in the determination are usually existent in his repertoire. Obviously, programs involving many discrimination frames require a multitude of examples and models. We are in the very early stages of exploring the problem of optimizing the examples we use.

The study referred to in the title of this paper pertains to research that two of us at the Center for Research on Learning and Teaching are involved in — research funded by the Office of Education.[1] My co-investigator, Martin Rothenberg, had developed a preliminary discrimination program designed to teach students of linguistics and communication sciences how to read visible speech patterns. Figure 2 on the handout is an example of such a pattern produced by a speech sound being fed through an apparatus which graphically portrays the duration, the frequencies and, by degree of blackness, the intensity of sound. After the program was developed we considered the possibility of teaching speech sounds with abstracted rather than real patterns. These patterns, made of art materials, would present only the relevant dimensions of the visible stimulus. The pattern labeled "abstract" in Figure 2 is an example. At the grossest level of questioning we were concerned with the relative merits of teaching via abstraction and via real patterns. Classically, this might be termed a transfer of training problem. It can also be viewed as a specific case of the general question: what are the criteria for selecting examples in a program? Rather than describe in detail the research underway, which incidentally is only in the first stages, this paper will attempt to explicate the general question that the research addresses itself to, however primitively.

Assume that we were teaching the concept of jar as opposed to the concept bottle as opposed to non-jar and non-bottle. Obviously there are many dimensions along which the concept bottle might be taught; for example, there are glass bottles, and ceramic bottles; bottles come in various colors and so do jars; bottles tend to have corks or crimped metal caps while jars tend to have screw-on caps. And so on. Let us take a single distinguishing dimension: size of mouth. In Figure 3 of the handout an attempt is made to show visually an assumed distribution in regard to "mouth width" of the real world population of objects called jar and those called bottle. On the right the normal distribution represents the frequency in the real world with which objects called jars having a particular mouth width are found. The typical jar, at least typical in terms of mouth width, would of course be found at the center of this curve. On the left side of the figure is the distribution of real bottles in terms of size of mouth. We could erect similar distributions for any particular dimensions of the two concepts.

---

[1] The research referred to is being conducted under contract No. OE-5-10-196 with the Office of Education; Project S-139.

These distributions are quite analogous to those we could develop in the visible speech project. The same vowel or consonant when spoken by different speakers, or in different context, or, even, repeatedly in the same context by the same speaker, varies somewhat and, to complicate things further, varies differently in different dimensions. So that we might easily take the distribution shown and label it for one vowel sound "ah" and the other vowel sound "eh", the horizontal axis would be labeled: "location of first formant in cycles per second."

Now, if we were selecting examples to teach these two concepts we might consider the distribution on the left and the distribution on the right, and compare them with the distribution of examples we originally had planned to use when teaching. (Incidentally, we should note that these distributions not only do not exhaust the many dimensions of the concepts indicated, but also do not indicate the distributions of non-bottles and non-jars which, of course, we would want to consider when selecting a second set of negative examples of the concept jar.) Comparing the real world distributions of items or examples with our repertoire of examples chosen to teach with, we might find some major differences, and typically, looking at programs and other teaching materials, one does find that the real world distributions of examples differ markedly from those used in teaching.

For the sake of convenience we might have chosen speech samples from the same speaker who represents typical, or quite atypical, articulation.

Thus, a male speaker will most always have a lower pitch. None of his speech records will precisely duplicate those of a female speaker.

We might find, for example, that either all or none of the teaching examples come from the most frequently found instance in the real population. Suppose that all of the examples in teaching came from the typical case —— the mean or average example. Figure 4, curve A shows such a case. The student in this program would hardly be ready to deal with the less frequent examples of the concept found in the real world. With no training in generalization he might either fail to generalize to the new cases or he might overgeneralize incorrectly to cases which are not truly in the distribution. On the other hand, if none of our examples in teaching are drawn from typical real life instances (Figure 4B) then most of our training effort and most of the teaching time has been in the presence of examples which the student is unlikely to meet at the expense of presenting a more realistic context.

Let us return to Figure 3 and look at the shaded area, the overlapping area of the two curves. I suppose this is graphic presentation of ambiguity. The real world is full of such examples. Take the phrase bench-table which indicates something that is, indeed, part table and part bench and tells us by its name that we might use that piece of furniture either as a bench or a table. The blending in the word brunch indicates that the meal is ambiguous in relation to either lunch or breakfast.

Someone coming new upon the academic world would suspect that there was, indeed, no ambiguity in the world for we rarely if ever deal with it in teaching. Rarely does a frame in a program deal with the problem of overlapping concepts of ambiguities. Were we teaching the concepts jar and bottle we might well want to indicate to the student the degree to which ambiguity occurs in the real world; teach him that there is overlap, and indicate how he might deal with it. Perhaps we would teach him to call these objects 'jottles', or he might learn to use the phrase, rarely heard in academia, "I simply can't answer that one."

Programs alone do not show this deficiency. Presentations of "typical" patients in clinical demonstrations to medical students usually involve single examples, chosen for reasons of availability or expediency. Teaching diagnosis obviously should involve the careful consideration of examples and explicit teaching of their typicality. Illustrations in textbooks not only present too few examples but often only typical, and sometimes only atypical, examples.

I would suggest from my own experience that in teaching most often we choose instances from opposite ends of the two distributions of the concepts being contrasted. These examples supposedly are clearest and emphasize only relevant features. However, they not only occur infrequently in the real world but also fail to produce the clearcut and precise discriminations that would be produced if similar contrasts were used.

If there were no other considerations in choosing examples the population of examples in teaching material probably should be distributed in a way similar to the distribution in real life in order to produce maximal transfer. No matter how intriguing the unique instance is to the teacher, no matter how helpful the teacher tries to be by purifying and abstracting an example, one step in optimizing examples would be to parallel the world in which the student will move. As in statistics, a measure of central tendency, in this instance using typical examples, is probably useless and may even lead to false conclusions when unaccompanied by an indication of the range of variability.

However, there are other considerations when deciding on examples. So far I have suggested that a formal correspondence between examples in the program and in real life is a desideratum. It is only a beginning point. Functionally, in terms of pedagogical effectiveness, we may have to include many more examples in, say, an area of ambiguity in order to produce a clear discrimination than we need use when teaching a discrimination between extreme cases. In short a formal similarity of example distributions (real-world and program) may not produce in the student the ability to make the correct real-world discriminations.

The discussion has been limited to examples drawn directly from real life. There is not time for a detailed exposition of problems concerned with the use of abstractions. Yet some mention must be made of it. Educational materials include a great amount of abstract "examples."[2] Drawings, not photographs, are commonly found in chapters of anatomy, physiology, engineering or biology textbooks. Emphasized, purified pronunciation, not true native speech, is typical in a language laboratory. Teaching stimuli are selected out and presented free from the noisy environment of real life. Such a distillation of a complex stimulus situation may lead to quicker acquisition (though some evidence from our research suggests it does not; other evidence suggests it does). Yet it may also lead to less transfer to the real life situation. (Again, some of our data supports and some questions this intuition.) A case can be made intuitively either for or against abstraction. What is needed is more empirical evidence defining for us those areas in which abstract examples are more efficient and effective than real ones.

One expects that, from an engineering standpoint, the optimal mixture of real and abstract examples will have to be determined for each particular teaching problem. The selections of examples must include considerations I have not dealt with. Obviously an additional one is the particular point of introduction in the program for a specific example or contrasting pair.

In summary, a general rule for improving programs might be "use more examples;" the use of examples is probably a first step in developing conceptual behavior in the student. Yet the simplicity of the rule belies the complexity involved in its application. If examples are to be used one must consider the criteria for choosing them.

---

[2] In the case of pure abstraction we are probably dealing with a rule rather than an example; most abstractions, however, are refined representations of real examples.

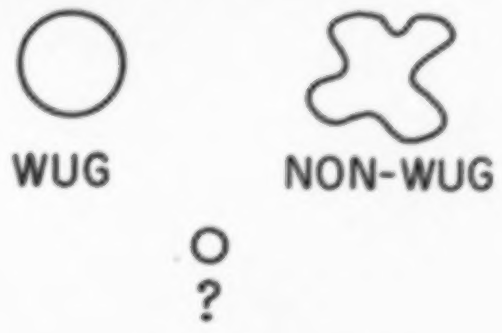


Figure 1

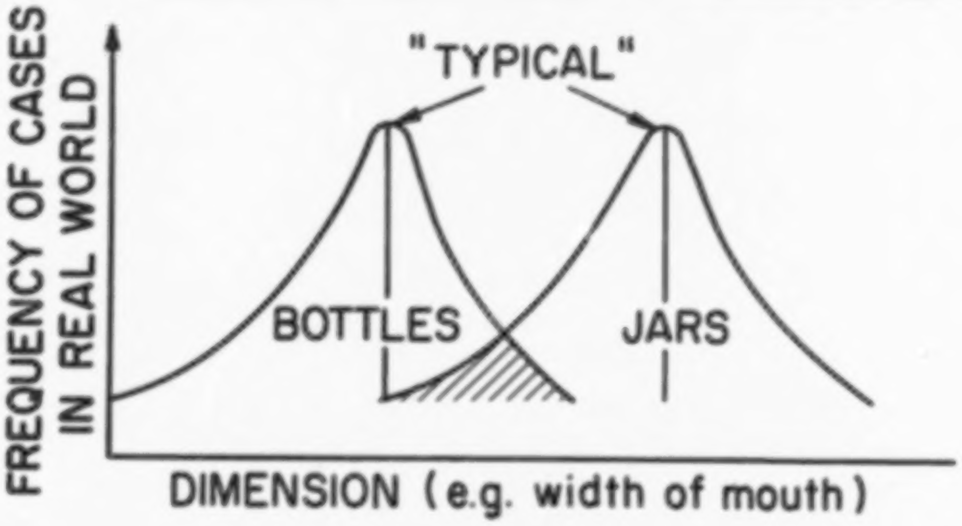


Figure 3

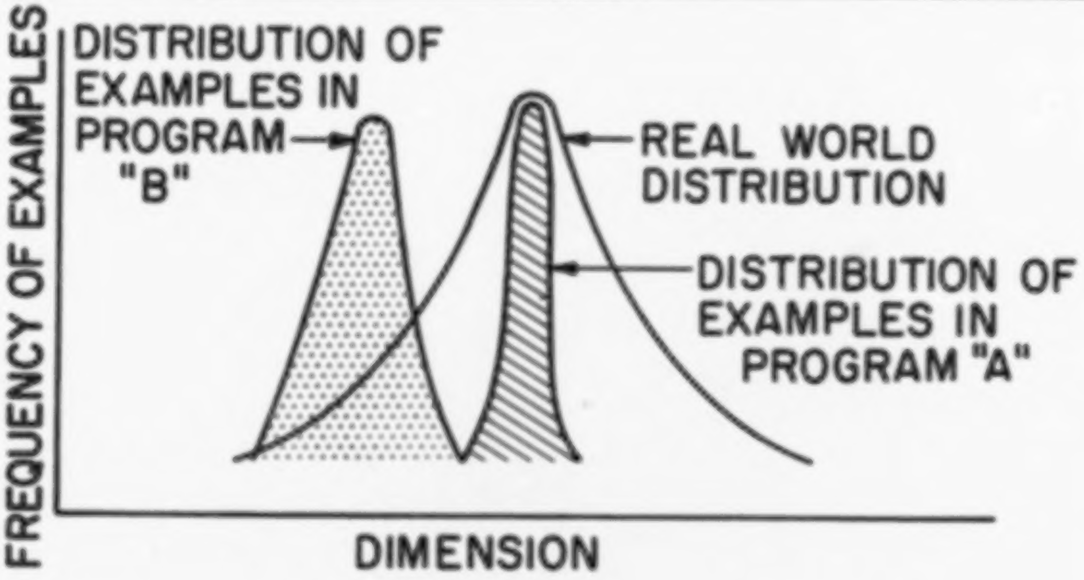


Figure 4